LYTIC ENZYMES AND SPORE SURFACE ANTIGEN FOR DETECTION AND TREATMENT OF *BACILLUS ANTHRACIS* BACTERIA AND SPORES

[01] This application claims the benefit of U.S. Patent Application Number 60/555,916, "Lytic Enzymes and Spore Surface Antigen for Detection and Treatment of *Bacillus Anthracis* Bacteria and Spores" (Fischetti, et. al.), filed March 24, 2004.

FIELD OF THE INVENTION

[02] The present invention relates to the identification and characterization of environmental bacteriophages infecting *Bacillus anthracis*. Specifically, the invention relates to certain isolated sequences for the gamma (γ)-phage and the W-phage of *B. anthracis*, nucleic acids of each genome, nucleic acids comprising nucleotide sequences of open reading frames (ORF's) of its genome, and polypeptides encoded by the nucleic acids.

BACKGROUND

[03] Anthrax is a disease is believed to be caused by the spore-forming bacterium, *Bacillus anthracis*, a bacterium that is readily found in soil. *B. anthracis* is believed to primarily cause disease in plant-eating animals. Though infrequent, when humans do become infected, they usually acquire the bacterium from contact with infected animals, animal hides or hair, or animal feces. The human disease has a relatively short incubation period (less than a week) and usually progresses rapidly to a fatal outcome.

[04] In humans, anthrax may occur in three different forms: coetaneous anthrax, gastrointestinal anthrax and inhalation anthrax. Coetaneous anthrax, the most common form in humans, is usually acquired when the bacterium, or spores of the bacterium, enter the body through an abrasion or cut on the skin. The bacteria multiply at the site of the abrasion, cause a local edema, and a series of skin lesions--papule, vesicle, pustule and necrotic ulcer--are sequentially produced. Lymph nodes nearby the site are eventually infected by the bacteria and, in cases where the organisms then enter the bloodstream (20% of cases), the

disease is often fatal. Gastrointestinal anthrax is caused by eating contaminated meat. Initial symptoms include nausea, vomiting and fever. Later, infected individuals present with abdominal pain, severe diarrhea and vomiting of blood. This type of anthrax is fatal in 25% to 60% of cases. Inhalation anthrax (also called woolsorters' disease) is acquired through inhalation of the bacteria or spores. Initial symptoms are similar to those of a common cold. Symptoms then worsen and these individuals present with high fever, chest pain and breathing problems. The infection normally progresses systemically and produces a hemorrhagic pathology. Inhalation anthrax is fatal in almost 100% of cases.

[05] Coetaneous anthrax is acquired via injured skin or membranes, entry sites where the spore germinate into vegetative cells. Proliferation of vegetative cells results in gelatinous edema. Alternatively, inhalation of the spores results in high fever and chest pain. Both types may be fatal unless the invasive aspect of the infection may be intercepted.

[06] *B. anthracis* is believed to possess two major virulence components. The first virulence component is a polysaccharide capsule which contains poly-D-glutamate polypeptide. The poly-D-glutamate capsule is not itself toxic but plays an important role in protecting the bacterium against anti-bacterial components of serum and phagocytic engulfment. As the *B. anthracis* bacterium multiplies in the host, it produces a secreted toxin which is the second virulence component of the organism. This anthrax toxin mediates symptoms of the disease in humans.

[07] The anthrax toxin is believed to comprise three distinct proteins encoded by the bacterium: protective antigen (PA), lethal factor (LF) and edema factor (EF). PA is the component of the anthrax toxin that is believed to bind to host cells using an unidentified cell-surface receptor. Once it binds to cell surfaces, EF or LF may subsequently interact with the bound PA. The complexes are then internalized by the host cell with significant effects. EF is an adenylate cyclase which causes deregulation of cellular physiology, resulting in edema. LF is a metalloprotease that cleaves specific signal transduction molecules within the cell (MAP kinase isoforms), causing deregulation of said pathways, and cell death. Injection of PA, LF or EF alone, or LF in combination with EF, into experimental

animals produces no effects. However, injection of PA plus EF produces edema. Injection of PA plus LF is lethal, as is injection of PA plus EF plus LF.

[08] As an acute, febrile disease of virtually all warm-blooded animals, including man, anthrax can be used in biological weapons (BW). For example, ten grams of anthrax spore may kill as many people as a ton of the chemical warfare agent, sarin. Terrorists have included dry spores in letters. Biological weapons of mass destruction have been developed that contain large quantities of anthrax spores for release over enemy territory. Once released, spores may contaminate a wide geographical area, infecting nearly all susceptible mammals. Due to the spore's resistance to heat and dry conditions, contaminated land may remain a danger for years. In view of the serious threat posed by the disease, effective diagnostic tools are needed to assist in prevention and control of natural and man-made outbreaks. Due to the highly lethal nature of anthrax and BW agents in general, there is great need for the development of sensitive and rapid BW agent detection. Current detection technology for biological warfare agents have traditionally relied on time-consuming laboratory analysis or onset of illness among people exposed to the BW agent.

[09] One promising approach to the detection and treatment of *B. anthracis* is the use of bacteriophage lysins as bacteriolytic agents. Bacteriophages specific for *B. anthracis* and related *B. cereus* bacteria strains may be isolated and used to detect and treat these bacteria. Bacteriophages near *B. anthracis* spores during spore germination may be used to infect and lyse the bacteria. A variety of phage-based bacterial therapies have been reviewed. D.H. Duckworth, P.A. Gulig, "Bacteriophages: Potential treatment for bacterial infections," *BioDrugs*, 16(1), 57-62 (2002). There are various environmental bacteriophages present in soils that may infect and lyse *B. anthracis* under controlled conditions. H..W. Ackermann, et al., "New Bacillus bacteriophage species," *Archives of* Virology, 135(3-4), 333-344 (1994); H.W. Ackerman, M.S. Dubrow, *Viruses of prokaryotes: General properties of bacteriophages*, Boca Raton, Fl, CRC Press, Inc. (1989);

[10] A bacterial lysin called PlyG, from bacteriophage-γ of *B. anthraci,* has been shown to lyse vegetative *B.* anthracis cells and is useful in promising methods for treatment of anthrax. R. Schuch, D. Nelson, V. Fischetti, "A bacteriolytic agent that detects and kills *Bacillus anthracis*," *Nature* 418, 884-889 (2002), incorporated herein by reference. A nucleotide sequence encoding PlyG is disclosed in GenBank accession #AF536823 and has a molecular mass of about 27,000. PlyG has been shown to control anthrax disease in mice, and to bind to vegetative cells. However, PlyG has no means to replicate itself in the presence of host bacteria. Methods and composition for the treatment of a variety of bacterial infections using a phage associated lytic enzyme specific for the invasive bacteria and an appropriate carrier for delivering the lytic enzyme into a patient are discussed in the following U.S. Patents issued to Fischetti et al.: 5,604,109; 5,985,271; 6,056,954; 6,056,955 6,248,324; 6,254,866; and 6,264,945, all incorporated herein by reference. Effective treatment of 14 of 24 virulent *B. anthracis* strains by phage based methods has been reported in a preliminary study done at Johns Hopkins University Applied Physics Laboratory. Michael Walter, Ph.D., "Efficacy and Durability of *Bacilus anthracis* Bacteriophages Used Against Spores," *Journal of Environmental Health*, July/August 2003, 9-15.

[11] Bacteriophages for *B. anthracis* may be isolated from the environment. For instance, Walter et al. report the isolation of Phages Nk, DB and MH for *B. anthracis* in topsoil. Walter, MH, Baker, DD, "Three *Bacillus anthracis* bacteriophages from topsoil," Curr Microbiol. 2003 Jul; 47(1): 55-58. Further bacteriophages useful for detection and treatment of *B. anthracis* are reported herein. The W and γ environmental bacteriophages of *B. anthracis* have been identified in topsoil, but the isolation of the polynucleotide and the identification of open reading frames coding for various polypeptides therein were unknown. E.W. McCloy, "Studies of a lysogenic *Bacillus* strain. I.. A bacteriophage specific for *Bacillus anthracis*," *Journal of Hygiene*, 49(1), 114-125 (1951); E.R. Brown, W.B. Cherry, "Specific identification of *Bacillus anthracis* by means of a variant bacteriophage," *Journal of Infectious Diseases,* 96(1), 34-39 (1955).

[12] The direct introduction of bacteriophages into an animal to prevent or fight diseases has certain drawbacks. Specifically, both the bacteria and the phage have to be in the correct and synchronized growth cycles for the phage to attach. Additionally, there are preferably the right number of phages to attach to the bacteria; if there are too many or too few phages, there will be either no attachment or no production of the lysing enzyme. The phage is preferably active enough to be effective. The phages may also be inhibited by many things including bacterial debris from the organism it is going to attack. Further complicating the direct use of a bacteriophage to treat bacterial infections is the possibility of immunological reactions within the subject being treated, potentially rendering the phage non-functional. The ability of bacteriophages to lyse and kill target bacterial may also be decreased by sunlight, UV light, desiccation or other conditions encountered during storage or use of a phage-containing therapeutic agent. Therefore, the potential effectiveness of any given bacteriophage against a target bacteria depends on the conditions under which the phage is deployed against the target bacteria. Studying the structure of phages and their efficacy against target bacteria in various conditions are useful in developing therapeutic methods for treating and preventing disease caused by target bacteria. Investigations of the structure and function of phages may also relate to diagnostic methods for detecting target bacteria and spores, such as those of *B. anthracis*. Many environmental conditions that may alter the effectiveness of a phage, such as phage W and phage-γ, against a *B. anthracis* or related target bacteria. The isolation and analysis of the phage polynucleotide sequences, and associated polypeptide sequences, of these and other phages are needed to relate to effective methods for prevention, treatment and diagnosis of *B. anthracis* bacteria and spores.

OVERVIEW

[13] Two bacteriophages of *B. anthracis*, bacteriophage gamma (γ) and bacteriophage W, can be isolated. Applicants have isolated and characterized

various bacteriophages active against *B. anthracis*. The γ and W bacteriophages for *B. anthracis*, the nucleic acid sequence of these bacteriophage genomes, portions of the nucleic acid sequence of the bacteriophage genome (*e.g.*, a portion containing an open reading frame), and proteins encoded by the nucleic acid sequences, as well as nucleic acid constructs comprising portions of the nucleic acid sequence of the bacteriophage genome, and host cells comprising such nucleic acid constructs are provided herein.

[14] More particularly, in some embodiments, the invention relates to certain nucleic acids of the genome of bacteriophages γ and W, as well as to the nucleic acids of portions of the genome of bacteriophages γ and W; to isolated nucleic acid molecules containing a nucleotide sequence of an open reading frame (or more than one open reading frame) of the genomes of bacteriophages γ and W; to isolated nucleic acid molecules encoding a polypeptide obtainable from bacteriophages γ and W or an active derivative or fragment of the polypeptide (e.g., a DNA polymerase, such as a DNA polymerase lacking exonuclease domains; a 3'-5' exonuclease, such as a 3'-5' exonuclease lacking DNA polymerase domain; a 5'-3' exonuclease (RNase H); a DNA helicase; or an RNA ligase); to DNA constructs containing the isolated nucleic acid molecule operatively linked to a regulatory sequence; and also to host cells comprising the DNA constructs. The invention further relates to isolated polypeptides encoded by these nucleic acids, as well as active derivatives or fragments of the polypeptides.

[15] In particular embodiments, the present invention relates to an isolated nucleic acid sequences that are at least 60%, 70%, 80%, 90%, 95%, 97%, 98-100% or 100% identical to a polynucleotide sequences encoded by SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112 or SEQ ID NO:113 and to a polynucleotide sequence encoding a polypeptide selected from SEQ ID NO:3- SEQ ID NO:109. In other embodiments, the polynucleotide of the invention is an isolated nucleic acid consisting of the sequence of SEQ ID NO:1 or SEQ ID NO:2, and an open reading frame (ORF) portion therein as identified in Table 1 or Table 2 below. The invention relates to an isolated nucleic acid

sequence of SEQ ID NO:1, an open reading frame of SEQ ID NO:1 set forth in Table 1, SEQ ID NO:2, or an open reading frame of SEQ ID NO:2 set forth in Table 2, with up to 5, 10, 20, 30, 40, 50, 60, 80, 100 or more conservative nucleic acid substitutions. Further provided are nucleic acid sequences of SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112 or SEQ ID NO:113 with up to 5, 10 or 20 conservative nucleic acid substitutions. The invention also relates to an isolated nucleic acid molecule comprising 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more contiguous nucleotides of SEQ ID NO:1 or SEQ ID NO:2. Other embodiments relate to an isolated nucleic acid molecule comprising contiguous nucleotides of an open reading frame from SEQ ID NO:1 or SEQ ID NO:2. Still other embodiments relate to a DNA construct comprising an isolated nucleic acid molecule comprising the nucleotide sequence of an open reading frame SEQ ID NO:1 or SEQ ID NO:2, operatively linked to a regulatory sequence, or the nucleic acid sequences of SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112 or SEQ ID NO:113.

[16] The invention further relates to a polypeptide comprising a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109. Some embodiments relate to a purified polypeptide, the amino acid sequence of which comprises a sequence at least 60%, 70%, 80%, 90%, 95%, 97%, 98-100% or 100% identical to a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109. Also provided is an isolated nucleic acid sequence encoding a polypeptide comprising the amino acid sequence set forth in a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109, with up to 5, 10, 20, 30, 40, 50, 60, 80, 100 or more conservative amino acid substitutions. The invention also relates to a purified polypeptide, the amino acid sequence of which consists of a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109. Other embodiments of the invention relate to a purified polypeptide, the amino acid sequence of which is encoded by an open reading frame from SEQ ID NO:1 or SEQ ID NO:2.

[17] The invention relates to polypeptides encoded by SEQ ID NO:1 or SEQ ID NO:2 that are able to infect *B. anthracis* or RSVF1 bacteria. Particular embodiments relate to polypeptide sequences that infect *B. anthracis* or RSVF1 to

a greater extent than other *B. cereus* bacteria. For example, some polypeptides of the invention may bind to *B. cereus* bacteria other than *B. anthracis* or RSVF1 at less than 100 plaque forming units/ml (PFU/ml), or even less than 10 PFU/ml, or less than 1 PFU/ml. In one embodiment, the invention relates to polypeptides encoded by SEQ ID NO:1 or SEQ ID NO:2 that are able to specifically bind to *B. anthracis* or RSVF1. The nucleic acid may encode one or more polypeptides that are able to infect *B. anthracis*. The nucleic acid may also encode one or more polypeptides that are able to bind to the surface of *B. anthracis*. The nucleic acid may also encode one or more polypeptides that exhibit fosfomycin resistance. The nucleic acid may encode one or more polypeptides that are spore surface antigens of *B. anthracis*.

[18] In one embodiment, the invention relates to a polypeptide encoded by the ORF 14 portion of SEQ ID NO:1, the polypeptide of SEQ ID NO:29, the polypeptide encoded by the ORF 14 portion of SEQ ID NO:2, or the polypeptide of SEQ ID NO:30, wherein the polypeptide is able to bind to the surface of *B. anthracis*. In another embodiment, the invention relates to a polypeptide encoded by the ORF17 portion of SEQ ID NO:1, the polypeptide of SEQ ID NO:35, the polypeptide encoded by the ORF 17 portion of SEQ ID NO:2, or the polypeptide of SEQ ID NO:36, wherein the polypeptide kills *B. anthracis*. In yet another embodiment, the invention relates to a polypeptide encoded by the ORF 41 portion of SEQ ID NO:1, or the polypeptide of SEQ ID NO:83, wherein the polypeptide exhibits Fosfomycin resistance. In further embodiments, the invention relates to the polypeptide encoded by the ORF 39 portion of SEQ ID NO:2, or the polypeptide of SEQ ID NO:82, wherein the polypeptide is a surface antigen of *B. anthracis*.

[19] Further provided are isolated nucleic acids that hybridize under high stringency conditions to the sequence of SEQ ID NO:1, SEQ ID NO:2, or open reading frame portions thereof as detailed in Table 1 and Table 2. In one embodiment, the invention relates to an isolated nucleic acid that hybridizes under high stringency conditions to a nucleic acid encodes a polypeptide that comprises a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109.

[20] Also provided is an isolated nucleic acid that hybridizes under high stringency conditions to the sequence of the ORF 14 from SEQ ID NO:1, or the ORF 14 from SEQ ID NO:2, wherein said nucleic acid encodes a polypeptide that is able to bind to the surface of *B. anthracis*. Further provided is an isolated nucleic acid that hybridizes under high stringency conditions to the sequence of the ORF 17 from SEQ ID NO:1, or the ORF 17 from SEQ ID NO:2, wherein said nucleic acid encodes a polypeptide that is kills *B. anthracis*. Further provided is an isolated nucleic acid that hybridizes under high stringency conditions to the sequence of the ORF 41 from SEQ ID NO:1, wherein said nucleic acid encodes a polypeptide exhibits Fosfomycin resistance. Also provided is an isolated nucleic acid that hybridizes under high stringency conditions to the sequence of the ORF 39 from SEQ ID NO:2, wherein said nucleic acid is a spore surface antigen of *B. anthracis*.

[21] Further provided are expression vectors comprising the nucleic acid sequence associated with ORF 14 from SEQ ID NO:1, the ORF 14 from SEQ ID NO:2, the ORF 17 from SEQ ID NO:1, the ORF 17 from SEQ ID NO:2, the ORF 41 from SEQ ID NO:1 or the ORF 39 from SEQ ID NO:2, operably associated with a promoter, and associated host cells comprising these vectors. Further provided are methods for preparing a polypeptide, each method comprising the step of culturing the host cell comprising the nucleotide sequence associated with ORF 14 from SEQ ID NO:1, the ORF 14 from SEQ ID NO:2, the ORF 17 from SEQ ID NO:1, the ORF 17 from SEQ ID NO:2, the ORF 41 from SEQ ID NO:1 or the ORF 39 from SEQ ID NO:2, under conditions that permit expression of the polypeptide from the expression vector, and isolating the polypeptide from the host cell. The invention also relates to an isolated nucleic acid comprising a sequence that hybridizes under high stringency conditions to a hybridization probe, the nucleotide sequence of which comprises or consists of ORF 14 from SEQ ID NO:1, the ORF 14 from SEQ ID NO:2, the ORF 17 from SEQ ID NO:1, the ORF 17 from SEQ ID NO:2, the ORF 41 from SEQ ID NO:1 or the ORF 39 from SEQ ID NO:2, or the complements thereof. Further provided is an isolated nucleic acid comprising a sequence that hybridizes under high stringency

conditions to a hybridization probe, the nucleotide sequence of which encodes the protein of a polypeptide sequence encoded by the ORF 14 from SEQ ID NO:1, the ORF 14 from SEQ ID NO:2, the ORF 17 from SEQ ID NO:1, the ORF 17 from SEQ ID NO:2, the ORF 41 from SEQ ID NO:1 or the ORF 39 from SEQ ID NO:2, or the nucleotide sequence of which encodes the protein encoded by these ORFs.

[22] Other embodiments of the instant invention include an isolated nucleic acid comprising a sequence that hybridizes under high stringency conditions to a hybridization probe, the nucleotide sequence of which consists of an open reading frame from SEQ ID NO:1 from Table 1, an open reading frame from SEQ ID NO:2 from Table 2, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112 or SEQ ID NO:113 or the complements thereof. The invention also relates to an isolated nucleic acid comprising a sequence that hybridizes under high stringency conditions to a hybridization probe, the nucleotide sequence of which encodes the protein of a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109.

[23] The invention further relates to an expression vector comprising the nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:2 or an open reading frame thereof as noted in Table 1 or Table 2, operably associated with a promoter, or a host cell comprising said vector. The invention also relates to an isolated nucleic acid comprising a sequence that encodes a protein of a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109, operably associated with a promoter, or a host cell comprising said vector. .

[24] The invention also relates to methods for preparing a polypeptide, the method comprising the step of culturing the host cell under conditions that permit expression of the polypeptide from the expression vector, and isolating the polypeptide from the host cell.

[25] The invention also relates to methods of screening for a compound that binds to a polypeptide, the method comprising: providing the nucleic acid of an open reading frame from SEQ ID NO:1 or SEQ ID NO:2, or an isolated nucleic acid comprising a sequence that encodes a protein of a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109, and introducing the nucleic acid

into a cell and allowing the cell to produce the polypeptide encoded by the nucleic acid, contacting a test compound with the polypeptide, and determining whether the test compound has bound to the polypeptide.

[26] The invention also relates to a method of screening for a compound that binds to a polypeptide, the method comprising: providing the nucleic acid encoding the polypeptide selected from the group consisting of: SEQ ID NO:3 – SEQ ID NO:109, introducing the nucleic acid into a cell and allowing the cell to produce the polypeptide encoded by the nucleic acid, contacting a test compound with the polypeptide, and determining whether the test compound has bound to the polypeptide.

BRIEF DESCRIPTION OF THE DRAWINGS

[27] **FIGURES. 1A-1B** (FIG. 1A, FIG. 1B) show the nucleotide (FIG. 1A; SEQ ID NO:1) and amino acid (FIG. 1B) sequences of *Bacillus anthracis* bacteriophage-γ.

[28] **FIGURES. 2A-2B** show the nucleotide (FIG. 1A; SEQ ID NO:2) and amino acid (FIG. 1B) sequences of *Bacillus anthracis* bacteriophage-W.

[29] **FIGURE 3A** is an electron microscope view of the page-γ. **FIGURE 3B** is an electron micrograph of page-γ particles adhered to cellular debris via the tips of the tail fibers. **FIGURE 3C** is an electron micrograph showing induction of unusual ring-shaped colonies of *B. cereus* strain ATCC 11950 with phage-W in the presence of fosfomyin. **FIGURE 3D** is an electron micrograph showing absence of ring-shaped colonies of *B. cereus* strain ATCC 11950 with phage-W in the absence of fosfomyin.

[30] **FIGURE 4A** and **FIGURE 4B** show analysis of the *Bacillus anthracis* genome (FIG. 4A) compared with other *Bacillus spp.* and *Clostridium* (FIG. 4B). **FIGURE 4C** is a graph showing results from the introduction of the pDG148::pg41 clone into RSVF1 has resulted in a 4-log increase in resistance to the antibiotic fosfomycin.

[31] **FIGURE 5A** shows a comparison of the genomic sequences of both γ and W by pair wise comparison. **FIGURE 5B** is a schematic of the *Bacillus anthracis* genome. **FIGURE 5C** is a schematic showing an alignment of the W phage.

[32] **FIGURE 6A** is a micrograph showing GFP-PlyG – binding of *Bacillus anthracis* in a whole cell manner. **FIGURE 6B** is a micrograph showing GFP-PlyG – binding of RSVF1 at only a polar positions. **FIGURE 6C** is a micrograph showing GFP-PlyG – binding of rare RSVF1 derivatives that bind in a whole cell fashion. **FIGURE 6D** is a micrograph showing GFP-Gp14 – whole cell binding with *Bacillus anthracis*. **FIGURE 6E** is a micrograph showing GFP-Gp14 – polar cell binding with *Bacillus anthracis*. **FIGURE 6F** is a micrograph showing *Bacillus anthracis* lysogenized with W becomes polar. **FIGURE 6G** and **FIGURE 6H** are micrographs showing fluorescence of GFP-PlyG binding to *Bacillus anthracis*.

[33] **FIGURE 7A** and **FIGURE 7B** are micrographs showing the effects of lysogeny with W on RSVF1 and *B. anthracis* showing rod shape formation. **FIGURE 7C, FIGURE 7D, FIGURE 7E,** and **FIGURE 7F** are electron micrographs showing spore appearance by SEM under various conditions as described below.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[34] The invention relates to the identification and characterization of an environmental bacteriophage infecting both *B. anthracis* and a transition state *B. cereus* strain, and thus establishing a means for genetic exchange between the two. Lysogeny of either organism exerts profound phenotypic changes and with *B. anthracis*, involves the acquisition of *B. cereus*-like features.

[35] A definition of terms used and their applicability to the disclosure are provided as follows:

[36] In this context of the embodiments, the term "lytic enzyme genetically coded for by a bacteriophage" means a polypeptide having at least some lytic activity against the host bacteria. The polypeptide has a sequence that

encompasses a native sequence of a lytic enzyme and variants thereof. The polypeptide may be isolated from a variety of sources, such as from phage, or prepared by recombinant or synthetic methods, such as those by Garcia et al. Every polypeptide has two domains. One domain is a choline binding portion at the carboxyl terminal side and the other domain is an amidase activity that acts upon amide bonds in the peptidoglycan at the amino terminal side. Generally speaking, a lytic enzyme according to the disclosure is between 25,000 and 35,000 daltons in molecular weight and comprises a single polypeptide chain; however, this may vary depending on the enzyme chain. The molecular weight most conveniently is determined by assay on denaturing sodium dodecyl sulfate gel electrophoresis and comparison with molecular weight markers.

[37] The term "purified" means that the biological material has been measurably increased in concentration by any purification process, including by not limited to, column chromatography, HPLC, precipitation, electrophoresis, etc., thereby partially, substantially or completely removing impurities such as precursors or other chemicals involved in preparing the material. Hence, material that is homogenous or substantially homogenous (e.g., yields a single protein signal in a separation procedure such as electrophoresis or chromatography) is included within the meanings of isolated and purified. Skilled artisans will appreciated that the amount of purification necessary will depend upon the use of the material. For example, compositions intended for administration to humans ordinarily may be highly purified in accordance with regulatory standards.

[38] "A native sequence phage associated lytic enzyme" is a polypeptide having the same amino acid sequence as an enzyme derived from nature. Such native sequence enzyme may be isolated from nature or may be produced by recombinant or synthetic means. The term "native sequence enzyme" specifically encompasses naturally occurring forms (e.g., alternatively spliced or modified forms) and naturally-occurring variants of the enzyme. In one embodiment of the disclosure, the native sequence enzyme is a mature or full-length polypeptide that is genetically coded for by a gene from a bacteriophage specific for *Bacillus anthracis*. Of course, a number of variants are possible and known, as

acknowledged in publications such as Lopez et al., Microbial Drug Resistance 3: 199-211 (1997); Garcia et al., Gene 86: 81-88 (1990); Garcia et al., Proc. Natl. Acad. Sci. USA 85: 914-918 (1988); Garcia et al., Proc. Natl. Acad. Sci. USA 85: 914-918 (1988); Garcia et al., Streptococcal Genetics (J.J. Ferretti and Curtis eds., 1987); Lopez et al., FEMS Microbiol. Lett. 100: 439-448 (1992); Romero et al., J. Bacteriol. 172: 5064-5070 (1990); Ronda et al., Eur. J. Biochem. 164: 621-624 (1987) and Sanchez et al., Gene 61: 13-19 (1987). The contents of each of these references, particularly the sequence listings and associated text that compares the sequences, including statements about sequence homologies, are specifically incorporated by reference in their entireties.

[39] "A variant sequence phage associated lytic enzyme" means a functionally active lytic enzyme genetically coded for by a bacteriophage specific for *Bacillus anthracis*, as defined below, having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or even at least 99.5% sequence identity with the amino acid or polynucleotide sequences shown below, or portions thereof. Of course a skilled artisan readily will recognize portions of this sequence that are associated with functionalities such as binding, and catalyzing a reaction. Polypeptide sequences and nucleic acids that encode these sequences are contemplated by some embodiments that comprise at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or more of each functional domain or open reading frame from the sequences provided herein. Such portions of the total sequence are very useful for diagnostics as well as therapeutics/prophylaxis. In fact, sequences as short as 5 amino acids long have utility as epitopic markers for the phage. More desirably, larger fragments or regions of protein having a size of at least 8, 9, 10, 12, 15 or 20 amino acids, and homologous sequences to these, have epitopic features and may be used either as small peptides or as sections of larger proteins according to embodiments. Nucleic acids corresponding to these sequences also are contemplated.

[40] Such phage associated lytic enzyme variants include, for instance, lytic enzyme polypeptides wherein one or more amino acid residues are added, or deleted at the N or C terminus of the sequences provided. In an embodiment one

or more amino acids are substituted, deleted, and/or added to any position(s) in the sequence, or sequence portion. Ordinarily, a phage associated lytic enzyme will have at least about (e.g. exactly) 50%, 55%, 60%, 65%, 70%, 75%, amino acid sequence identity with native phage associated lytic enzyme sequences, more preferably at least about (e.g. exactly) 80%, 85%, 90%, 95%, 97%, 98%, 99% or 99.5% amino acid sequence identity. In other embodiments a phage associated lytic enzyme variant will have at least about 50% (e.g. exactly 50%) , 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or even at least 99.5% amino acid sequence identity with the sequences provided, or portions thereof.

[41] A polypeptide or amino acid "selected from SEQ ID NO:3– SEQ ID NO:109" refers to a polypeptide sequence selected from the group consisting of: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12,SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22,SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32,SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42,SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52,SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62,SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72,SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82,SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92,SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102,SEQ

ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, and SEQ ID NO:109.

[42] "Percent amino acid sequence identity" with respect to the phage associated lytic enzyme sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the phage associated lytic enzyme sequence, after aligning the sequences in the same reading frame and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity may be achieved in various ways that are within the skill in the art, such as using publicly available computer software such as blast software. Those skilled in the art may determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the whole length of the sequences being compared.

[43] In each case, of course conservative amino acid substitutions also may be made simultaneously in determining percent amino acid sequence identity. For example, a 15 amino acid long region of protein may have 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% sequence homology with a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109, or portions thereof. At the same time, the 15 amino acid long region of the protein may also have up to 0.5%, 1%, 2%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 65%, 75%, or more amino acids replaced with conservative substitutions. Preferably the region will have fewer than 30%, 20%, 10% or even less conservative substitutions. The "percent amino acid sequence identity" calculation in such cases will be higher than the actual percent sequence identity when conservative amino acid substitutions have been made.

[44] "Percent nucleic acid sequence identity" with respect to the phage associated lytic enzyme sequences identified herein is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in the phage associated lytic enzyme sequence, after aligning the sequences and

introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity may be achieved in various ways that are within the scope of those skilled in the art, including but not limited to the use of publicly available computer software. Those skilled in the art may determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

[45] "Polypeptide" refers to a molecule comprised of amino acids which correspond to those encoded by a polynucleotide sequence which is naturally occurring. The polypeptide may include conservative substitutions wherein the naturally occurring amino acid is replaced by one having similar properties, where such conservative substitutions do not alter the function of the polypeptide (see, for example, Lewin "Genes V" Oxford University Press Chapter 1, pp. 9-13 1994).

[46] A "chimeric protein" or "fusion protein" comprises all or (preferably a biologically active) part of a polypeptide of the disclosure operably linked to a heterologous polypeptide. Chimeric proteins or peptides are produced, for example, by combining two or more proteins having two or more active sites. Chimeric protein and peptides may act independently on the same or different molecules, and hence have a potential to treat two or more different bacterial infections at the same time. Chimeric proteins and peptides also are used to treat a bacterial infection by cleaving the cell wall in more than one location.

[47] The term "operably linked" means that the polypeptide of the disclosure and the heterologous polypeptide are fused in-frame. The heterologous polypeptide may be fused to the N-terminus or C-terminus of the polypeptide of the disclosure. Chimeric proteins are produced enzymatically by chemical synthesis, or by recombinant DNA technology. A number of chimeric lytic enzymes have been produced and studied. Gene E-L, a chimeric lysis constructed from bacteriophages phi X174 and MS2 lysis proteins E and L, respectively, was subjected to internal deletions to create a series of new E-L clones with altered lysis or killing properties. The lytic activities of the parental genes E, L, E-L, and the internal truncated forms of E-L were investigated in this study to characterize

the different lysis mechanism, based on differences in the architecture of the different membranes spanning domains. Electron microscopy and release of marker enzymes for the cytoplasmic and periplasmic spaces revealed that two different lysis mechanisms may be distinguished depending on penetration of the proteins of either the inner membrane or the inner and outer membranes of the E. coli. FEMS Microbiol. Lett. 1998 Jul 1, 164(1); 159-67 (incorporated herein by reference).

[48] Isolated bacteriophages γ and W may be used in the study of the relationship between the bacteriophages and their host cells (e.g., *B. anthracis*, such as *B. anthracis* species ITI 378). Isolated bacteriophages γ and W may also be used as a vector to deliver nucleic acids to a host cell; that is, the bacteriophage may be modified to deliver nucleic acids comprising a gene from an organism other than the bacteriophage (a "foreign" gene). For example, nucleic acids encoding a polypeptide (e.g., an enzyme or pharmaceutical peptide) may be inserted into the genome of bacteriophages γ and W, using standard techniques. The resultant modified bacteriophage may be then used to infect host cells, and the protein encoded by the foreign nucleic acids may then be produced.

[49] Phage, or bacterial viruses, are major mediators of bacterial genetic diversity. They persist in bacterial populations by stably integrating into the host genome (lysogenic growth as a prophage form) and/or by freely replicating within a host (lytic growth). During such passage the phage genome may acquire, maintain, and transmit "foreign" DNA (obtained from other phage or the bacterial host) which serves to enhance fitness of the host. This foreign DNA may promote bacterial exploitation of animal tissues (resulting from exotoxins, colonization factors, serum resistance proteins, etc.), and it is likely to promote survival in other niches as well. Despite the increasingly well described role for phage in pathogen evolution, their place in the pathogenesis of *B. anthracis* is unclear. Since the *B. anthracis* pool is so genetically uniform, it is unlikely that phage drive the mergence of distinctly pathogenic strains, as is the case for other Gram-positive pathogens like *Streptococcus pyogene* and *Staphylococcus aureus*. The role may

rather be related to interactions (or a relationship) between *B. anthracis* and transition state *B. cereus*. Such possibility is based on studies from the 1940's and 50's showing that a lysogenic phage from the soil, called W, and a obligately lytic derivative thereof, called γ, infect both *B. anthracis* and the rare transition state *B. cereus* strains and thus may transmit information between the two. More recent studies suggest that several distinct naturally occurring and laboratory-induced *B. anthracis* phage may also infect certain *B. cereus* strains, which may have represented transition state isolates.

[50] Without being limited by theory, it is believed that *B. anthracis* is a genetically monomorphic variant of the otherwise highly polymorphic *B. cereus* lineage, which also includes *B. cereus* and *B. thuringiensis*. *B. anthracis* isolates recovered from diverse geographical locations or from present and past outbreaks are genetically distinguishable largely by molecular typing schemes that discriminate distinct and stable allelic states based on variations of tandem nucleotide-repeat elements in a few hypervariable loci. Several alternate analyses of genetic polymorphisms (multilocus enzyme electophoretic studies, for example) show a very close phylogenetic relationship between *B. anthracis* and a group of rare *B. cereus* "transition state" strains, possessing both *B. anthracis*- and *B cereus*-like qualities and that may be more readily recoverable from *B. anthracis* outbreak sites than is bona fide *B. anthracis*. The significance of this relationship to the ecology of anthrax is unclear. Currently, little is known regarding the fate of *B. anthracis* in the environment after host death, although it is held to involve stagnancy in the form of an absolutely dormant spore. Here, we report the identification and characterization of an environmental bacteriophage infecting both *B. anthracis* and a transition state *B. cereus* strain, and thus establishing a means for genetic exchange between the two. Lysogeny of either organism exerts profound phenotypic changes and with *B. anthracis*, involves the acquisition of *B. cereus*-like features.

[51] One embodiment of the invention relates to isolated γ or W bacteriophage. "Isolated" γ or W bacteriophage refers to bacteriophage that has

been separated, partially or totally, from its native environment (e.g., separated from *B. anthracis* host cells) ("native bacteriophage"), and also refers to bacteriophage that has been chemically synthesized or recombinantly produced ("recombinant bacteriophage"). A bacteriophage that has been "recombinantly produced" refers to a bacteriophage that has been manufactured using recombinant DNA technology, such as by inserting the bacteriophage genome into an appropriate host cell (e.g., by introducing the genome itself into a host cell, or by incorporating the genome into a vector, which is then introduced into the host cell).

Isolation and Preparation of Bacteriophages

[52] Bacteriophages γ and W may be produced by inoculating appropriate host cells with the bacteriophage. Representative host cells in which the bacteriophage may replicate include *B. anthracis*. The host cells may be cultured in a suitable medium (e.g., medium 162 for Thermus as described by Degryse et al., Arch. Microbiol. 11 7:189-196 (1978), with 1/10 buffer and with 1% NaCl). In addition, the host cells may be cultured under conditions suitable for replication of the bacteriophage. For example, in a preferred embodiment, the host cells may be cultured at a temperature of at least approximately 50°C In a more preferred embodiment, the host cells may be cultured at a temperature between about 50° C. and about 80°C. The bacteriophage may also be stored in a cell lysate at about 4°C.

NUCLEIC ACID SEQUENCES

[53] Another embodiment of the invention relates to isolated nucleic acid sequences obtainable from the genome of bacteriophages γ and W.

[54] The nucleic acid molecules of the invention may be "isolated;" as used herein, an "isolated" nucleic acid molecule or nucleotide sequence is intended to mean a nucleic acid molecule or nucleotide sequence which is not flanked by nucleotide sequences which normally (in nature) flank the gene or nucleotide sequence (as in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (e.g., as in an RNA library). For

example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstance, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Thus, an isolated nucleic acid molecule or nucleotide sequence may include a nucleic acid molecule or nucleotide sequence which is synthesized chemically or by recombinant means. Therefore, recombinant DNA contained in a vector may be included in the definition of "isolated" as used herein. Also, isolated nucleotide sequences include recombinant DNA molecules in heterologous organisms, as well as partially or substantially purified DNA molecules in solution. In vivo and in vitro RNA transcripts of the DNA molecules of the present invention may be also encompassed by "isolated" nucleotide sequences.

[55] The present invention also relates to nucleotide sequences which may be not necessarily found in nature but that encode the polypeptides described below. Thus, DNA molecules which comprise a sequence which is different from the naturally-occurring nucleotide sequence but which, due to the degeneracy of the genetic code, encode the polypeptides described herein, such as SEQ ID NO:3 – SEQ ID NO:109, are also provided. Embodiments of the invention also encompass variations of the nucleotide sequences of the invention, such as those encoding active fragments or active derivatives of the polypeptides as described below. Such variations may be naturally-occurring, or non-naturally-occurring, such as those induced by various mutagens and mutagenic processes. Intended variations include, but are not limited to, addition, deletion and substitution of one or more nucleotides which may result in conservative or non-conservative amino acid changes, including additions and deletions. Preferably, the nucleotide or amino acid variations are silent or conserved; that is, they do not alter the characteristics or activity of the encoded polypeptide.

[56] The invention also relates to fragments of the isolated nucleic acid molecules described herein. The term "fragment" encompasses a portion of a

nucleotide sequence described that is from at least about 25 contiguous nucleotides to at least about 50 contiguous nucleotides or longer in length. These fragments may be useful as probes and also as primers. Particularly preferred primers and probes selectively hybridize to the nucleic acid molecule encoding the polypeptides described herein. For example, fragments that encode polypeptides that retain activity, as described below, may be particularly useful.

[57] The invention also relates to nucleic acid molecules that hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (e.g., nucleic acid molecules which specifically hybridize to a nucleotide sequence encoding polypeptides described herein, and, optionally, have an activity of the polypeptide). Hybridization probes may be oligonucleotides that may bind in a base-specific manner to a complementary strand of nucleic acid. Suitable probes include polypeptide nucleic acids, as described in (Nielsen et al., Science 254, 1497-1500 (1991)).

[58] These nucleic acid molecules may be detected and/or isolated by specific hybridization (e.g., under high stringency conditions). "Stringency conditions" for hybridization is a term of art which refers to the incubation and wash conditions, e.g., conditions of temperature and buffer concentration, which permit hybridization of a particular nucleic acid to a second nucleic acid; the first nucleic acid may be perfectly (i.e., 100%) complementary to the second, or the first and second may share some degree of complementarity which is less than perfect (e.g., 60%, 75%, 85%, 95%). For example, certain high stringency conditions may be used which distinguish perfectly complementary nucleic acids from those of less complementarity.

[59] "High stringency conditions," "moderate stringency conditions" and "low stringency conditions" for nucleic acid hybridizations are explained on pages 2.10.1-2.10.16 and pages 6.3.1-6 in Current Protocols in Molecular Biology (Ausubel, F. M. et al., "Current Protocols in Molecular Biology," John Wiley & Sons, (1998)) the teachings of which are hereby incorporated by reference. The exact conditions which determine the stringency of hybridization depend not only on ionic strength (e.g., 0.2.times.SSC, 0.1.times.SSC), temperature (e.g., room

temperature, 42°C., 68° C.) and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences and the frequency of occurrence of subsets of that sequence within other non-identical sequences. Thus, high, moderate or low stringency conditions may be determined empirically.

[60] By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions which will allow a given sequence to hybridize (e.g., selectively) with the most similar sequences in the sample may be determined.

[61] In one embodiment, PlyG may be used in the preparation of DNA, for example for hybridization studies. Using PlyG, DNA from *B. anthracis* can be rapidly and more gently extracted because of the specificity of PlyG for particular types of bacteria including *B. anthracis*. Accordingly, in this embodiment, less stringent hybridization conditions may be required to prepare DNA from bacteria that PlyG selectively acts upon than would otherwise be required in the absence of PlyG.

[62] Exemplary conditions are described in Krause, M. H. and S. A. Aaronson, Methods in Enzymology, 200:546-556 (1991). Also, in, Ausubel, et al., "Current Protocols in Molecular Biology," John Wiley & Sons, (1998), which describes the determination of washing conditions for moderate or low stringency conditions. Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, starting from the lowest temperature at which only homologous hybridization occurs, each °C. by which the final wash temperature is reduced (holding SSC concentration constant) allows an increase by 1% in the maximum extent of mismatching among the sequences that hybridize. Generally, doubling the concentration of SSC results in an increase in T.sub.m of .about.17° C. Using these guidelines, the washing temperature may be determined empirically for high, moderate or low stringency, depending on the level of mismatch sought.

[63] For example, a low stringency wash may comprise washing in a solution containing 0.2.times.SSC/0.1% SDS for 10 min at room temperature; a moderate stringency wash may comprise washing in a prewarmed solution (42° C.) solution containing 0.2.times.SSC/0.1% SDS for 15 min at 42° C.; and a high stringency wash may comprise washing in prewarmed (68° C.) solution containing 0.1.times.SSC/0.1%SDS for 15 min at 68° C. Furthermore, washes may be performed repeatedly or sequentially to obtain a desired result as known in the art.

[64] Equivalent conditions may be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleic acid molecule and the primer or probe used. Hybridizable nucleic acid molecules may be useful as probes and primers, e.g., for diagnostic applications.

[65] Examples of high stringency conditions may be selected from the group consisting of:

[66] (a) 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50°C;

[67] (b) 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C; and

[68] (c) 50% formamide, 5 x SSC (0.75 M sodium chloride, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% sodium dodecyl sulphate, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC (0.75 M sodium chloride, 0.075 M sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash consisting of 0.1 x SSC (0.75 M sodium chloride, 0.075 N sodium citrate) containing EDTA at 55°C.

[69] Such hybridizable nucleotide sequences may be useful as probes and primers for diagnostic applications. As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (e.g., in the presence of four

different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but are preferably sufficiently complementary to hybridize with a template. The term "primer site" refers to the area of the target DNA to which a primer hybridizes. The term "primer pair" refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

[70] As described herein, the genome of bacteriophages γ and W have been sequenced. The polynucleotide sequence of bacteriophage γ is set forth in **FIG. 1A** (SEQ ID NO:1), and corresponding polypeptide sequences for open reading frames of SEQ ID NO:1 are set forth in **FIG. 1B**. There are approximately 53 open reading frames (ORFs) in the polynucleotide sequence, as set forth in **Table 1**. The polynucleotide sequence of bacteriophage W is set forth in **FIG. 2A** (SEQ ID NO:2), and corresponding polypeptide sequences for open reading frames of SEQ ID NO:2 are set forth in **FIG. 2B**. There are approximately 54 open reading frames (ORFs) in the polynucleotide sequence of bacteriophage W, as set forth in **Table 2**. **Table 1** and **Table 2** relate to the locus of each ORF; the number of nucleotides in the ORF; the structure and function of various putative proteins encoded therein; the protein identified by a BLAST search as being the closest match to certain putative proteins; and other information relating to the ORFs.

[71] The invention thus relates to isolated nucleic acid sequence of the genome ("isolated genomic DNA") of the bacteriophages γ and W. The invention also relates to isolated nucleic acid sequence of the genome of bacteriophages γ and W. The invention additionally relates to isolated nucleic acid molecules comprising the nucleotide sequences of each of the ORFs described

above or fragments thereof, as well as nucleic acid molecules comprising nucleotide sequences of more than one of the ORFs described above or fragments of more than one of the ORFs. The nucleic acid molecules of the invention may be DNA, or may also be RNA, for example, mRNA. DNA molecules may be double-stranded or single-stranded; single stranded RNA or DNA may be either the coding, or sense, strand or the non-coding, or antisense, strand. Preferably, the nucleic acid molecule comprises at least about 100 nucleotides, or at least one ORF, or more preferably at least about 150 nucleotides, and even more preferably at least about 200 nucleotides. The nucleotide sequence may be only that which encodes at least a fragment of the amino acid sequence of a polypeptide; alternatively, the nucleotide sequence may include at least a fragment of a coding sequence along with additional non-coding sequences such as non-coding 3' and 5' sequences (including regulatory sequences, for example). Certain preferred nucleotide sequences of the invention may consist essentially of one of the ORFs and its flanking sequences. For example, in certain preferred embodiments, the nucleotide sequence comprises one of the following ORFs: ORF 14 of the g-phage (Gp14), ORF 14 of the W-phage (Wp14), and ORF 38 of the W-phage (Wp38).

Bacteriophage Gamma (γ)

[72] Analysis of the γ phage host range in **Table 1** demonstrated its specificity for *B. anthracis* and RSVF1. By electron microscopy (**FIG. 3A**), γ is morphologically similar to members of the *Siphoviridae* family of tailed phages (double-stranded DNA viruses of the order *Caudovirales*) consisting of a DNA-filled isometric, icosahedral head (90 nm in diameter) and a long non-contractile tail (260 nm in length) connected distally to a small plate and a fibrous tail extension (75 nm in length). Phage particles adhered to cellular debris via the tips of the tail fibers (**FIG. 3B**).

[73] A Fos resistance gene (ORF 14, or "Gp41") is found in γ and appears to be derived from a similar sequence in Ba prophage ϕ4537. This is based on 99% identity at the DNA level between *gp39*, *pg40* and *gp41* and sequences in a Ba prophage. The DNA surrounding this region of homology is quite divergent,

suggesting the acquisition of the island by γ through recombination with the ϕ4537 prophage. Since there is no obvious homology in W to have supported this insertion, it has likely arisen via a illegitimate mechanism.

[74] Gp41 encodes a fosfomycin resistance protein, or a glutathione transferase. Analysis of the Ba genome (**FIG. 4A**) shows ~35 such proteins, while other *Bacillus spp.* and *Clostridium* have 10-15 each (**FIG. 4B**). Only in the Ba genome is a copy prophage encoded and therefore mobile.

[75] We have cloned the Gp41 locus into an IPTG-inducible *Bacillus* protein expression vector (pDG148) and established a system for stably transforming RSVF1. Introduction of the pDG148::*pg41* clone into RSVF1 has resulted in a 4-log increase in resistance to the antibiotic fosfomycin (**FIG. 4C**). The MIC for fosfomycin was 62.5 μg ml^{-1} for RSVF1 and RSVF1/pDG148 and 500 μg ml^{-1} for RSVF1/pDG148::*gp41*.

Table 1. ORF's of Gamma (γ) Bacteriophage of *B. anthracis*

ORF	Frame	Position	Size	Matches (E value)*	Structure†	Function‡
1	3	54-539	18.5 (161)	Orf21 ϕ105 *B. subtilis* (5e-25), Phage terminase, small subunit (1e-6)	c-c domain	DNA packaging
2	2	536-2233	65.1 (565)	Orf22 ϕ105 *B. subtilis* (1e-151), Phage terminase, large subunit (2e-131)		DNA packaging
3	2	2249-3547	48.8 (432)	Gp3 ϕ3626 *C. perfringens* (e-112) Phage portal protein (9e-53)		Portal protein
4	3	3510-4130	23.7 (206)	Gp5 ϕ3626 *C. perfringens* (2e-49) *Caudovirales* prohead protease (9e-29)	c-c domain	Head maturation protease
5	2	4169-5347	44.2 (392)	Orf27 ϕ105 *B. subtilis* (4e-95) Phage capsid family (7e-72)	c-c domains	Major head protein
6	1	5365-5655	11.0 (96)	Gp7 ϕ3626 *C. perfringens* (2e-3) Phage QLRG family (1e-7)		
7	3	5652-5975	12.1 (107)	BA_4559 *B. anthracis* (6e-5) Bacteriophage head-tail adaptor (4e-9)	Pyrophosphatase domain	Putative head-tail adaptor
8	1	5968-6408	16.2 (146)	BA_4558 *B. anthracis* (2e-16) CRPp0301 (2e-5)		Uncharacterized protein
9	3	6405-6764	13.9 (119)	BA_4557 *B. anthracis* (4e-15) CRPp0346 (5e-8)		Uncharacterized protein
10	3	6765-7373	22.9 (202)	Chte_p_1640 *C. thermocellum* (9e-29) CRPp0161 (6e-11)	c-c domain	Major tail protein
11	1	7423-7740	11.8 (105)	BA_4555 *B. anthracis* (2e-3)		
12	3	7770-7946	7.0 (58)			
13	1	7963-11814	139.4 (1283)	BA_4552-BA_4554 *B. anthracis* (>6e-25), CRPp0381 (4e-40)	c-c domains	Tail protein
14	3	11829-13319	56.8 (496)	BA_4550 *B. anthracis* (e-153), CRPp0325 (2e-4)	c-c domain Pyridoxal-phosphate binding domain	Tail component
15	2	13316-17311	149.8 (1331)	BA_4578-BA_4579 *B. anthracis* (>5e-20), CRPp0329 (2e-36)	c-c domains	Similar to myosin heavy chain
16	1	17350-17775	15.0 (141)	BA_4545 *B. anthracis* (9e-57), Phage-related holin (3e-28)	3 TM domains	Holin, host lysis
17	3	17775-18476	26.3 (233)	BA_4545 *B. anthracis* (1e-112) Cell wall amidase (2e-43)		Lysin, host lysis
18	-1	19031-18534	18.0 (165)		c-c domain, 1 TM domain	May be a lipoprotein

ORF	Frame	Position	Size	Matches (E value)*	Structure†	Function‡
19	-3	19230-19018	8.3 (70)	BA_4541 *B. anthracis* (4e-13) Helix-turn-helix XRE-family (1e-4)	HTH domain only	
20	1	19414-19722	12.3 (102)	BA_4540 *B.anthracis* (2e-40)		
21	3	19719-19901	6.7 (60)	BA_4539 *B. anthracis* (3e-12)	2 TM domains	
22	3	19911-21200	49.3 (414)	BA_4538 *B. anthracis* (e-168) FtsK/SpoIIIE family: C-term. (1e-15)	c-c domain, P-loop (ATP/ GTP binding)	DNA translocation? Integration?
23	1	21178-21759	25.2 (175)	BC1920 *B. cereus* (2e-61)	c-c domain	
24	-3	22029-21793	8.6 (78)	BC1914 *B. cereus* (5e-18)		
25	2	21863-22090	8.4 (75)		2 TM domains	
26	1	22297-23160	33.3 (287)	BC4930 *B. cereus* (6e-11)	c-c domain	lysogeny?
27	1	23236-24681	56.4 (481)	Chte1631 *C. thermocellum* (2e-30) PinR, Site-specific recombinases (4e-20)	c-c domains	Integrase
28	-1	24812-24684	4.8 (39)	Helix-turn-helix XRE-family (0.01)	HTH domain fragment	Fragment of CI-like repressor?
29	3	24972-25199	8.8 (75)	BC2559 *B. cereus* (7e-5) Helix-turn-helix, Cro and CI (0.008)	HTH domain	Cro-like DNA binding role?
30	3	25212-25397	7.2 (61)	BA_4542 *B. anthracis* (7e-5)		
31	1	25642-26457	31.0 (271)	Orf6 *L. lactis* ϕbIL285 (2e-12), CRPp0355 (6e-17)	c-c domain	Anti-repressor
32	1	26524-27177	25.7 (217)	Orf16 *L. lactis* ϕbIL312 (4e-12)		
33	3	27306-28253	37.1 (315)	Orf11 *L. lactis* ϕrlt (1e-18) DnaA phage analogs	c-c domain AT-rich repeats	Phage replication; Origin of replication
34	3	28269-29180	34.9 (303)	DnaC DNA replication protein (3e-10) Ntp *Lactobacillus* ϕg1e (6e-29)	c-c domain, P-loop (ATP-GTP binding)	Phage replication
35	3	29199-29432	9.2 (77)			
36	1	29425-30171	28.1 (248)	BA_4585 *B. anthracis* (6e-41) FliA family of sigma factors (1e-20)		Transcriptional effects
37	3	30168-30644	19.0 (158)			
38	2	30704-31246	21.1 (180)	BA_5241 *B. anthracis* (2e-20)		
39	1	31270-31500	8.8 (76)	BA_4582 *B. anthracis* (1e-20)	2 TM domains (signal sequence)	Membrane protein
40	2	31607-32077	18.1 (156)	BA_4581 *B. anthracis* (2e-47)		
41	-1	32124-32513	15.5 (129)	BA_4580 *B. anthracis* (2e-58) Glyoxalase resistance protein (1e-7)		Fosfomycin resistance
42	3	32994-33164	6.5 (56)			
43	2	33353-33658	11.9 (101)		c-c domain	
44	3	33651-33890	9.3 (79)			
45	1	34327-34734	16.0 (135)	BC3700 *B. cereus* (1e-39)	c-c domains	
46	3	34854-35078	8.5 (74)		c-c domains	
47	3	35085-35306	8.2 (73)			30% identity to C-term. half of hamster IL-6
48	1	35311-35715	15.6 (134)			
49	3	35820-36011	7.2 (63)			
50	1	36031-36285	10.2 (84)		c-c domains	
51	1	36484-36675	7.3 (63)			
52	2	36656-36943	10.6 (95)	BA_4569 *B. anthracis* (5e-4)	c-c domains	

ORF	Frame	Position	Size	Matches (E value)*	Structure†	Function‡
53	1	36943-37326	15.5 (128)	Gp50 ϕ3626 *C. perfringens* (6e-13) MerA restriction endonuclease (2e-5)		endonuclease

*No entry indicates no significant homologies based on a protein-protein BLAST search. First line indicates the protein name, organism of origin, and BLAST E value for most significant hit. Second line indicates, if detected, the pfam conserved domain and E value or the cluster of related viral proteins (CRP) designation and E value.
†Indicates a significant protein structure or motif detected by bioinformatics analysis. Abbreviations are used: c-c domain/s, indicating the significant likelihood of one or more coiled-coil domains; TM, for transmembrane domain; and HTH, for helix-turn-helix.
‡Indicates putative function based on homologies detected with proteins of known function.

Bacteriophage W

[76] We also isolated the parental lysogenic phage, W. As part of a study of resistance to fosfomycin in *B. cereus* strains tested, ATCC 11950 produced unusual ring-shaped colonies when plated in the presence of fosfomyin (**FIG. 3C**), but not in the absence (**FIG. 3D**). The central clearing zone was found to be enriched for intact phage W particles, thus suggesting that the fosfomycin may have induced the phage from older colony members, which constitute the central portion of a colony. Much like γ, phage W infected both *B. anthracis* and RSVF1, and not other *B. cereus* or *B. thuringiensis* strains. Phage W was also morphologically identical to γ, confirming their close genetic relationship.

Table 2: ORF's of W-Bacteriophage of *B. anthracis*.

Wp	Frame	Position	Size	Matches (E value)*	Structure†	Function‡
1	3	54-539	18.5 (161)	Orf21 ϕ105 *B. subtilis* (5e-25), Phage terminase, small subunit (1e-6)	c-c domain	Terminase, small subunit
2	2	536-2233	65.1 (565)	Orf22 ϕ105 *B. subtilis* (1e-151), Phage terminase, large subunit (2e-131)		Terminase
3	2	2249-3547	48.8 (432)	Gp3 ϕ3626 *C. perfringens* (e-112) Phage portal protein (9e-53)		Portal protein
4	3	3510-4130	23.7 (206)	Gp5 ϕ3626 *C. perfringens* (2e-49) *Caudovirales* prohead protease (9e-29)	c-c domain	Head maturation protease
5	2	4169-5347	44.2 (392)	Orf27 ϕ105 *B. subtilis* (4e-95) Phage capsid family (7e-72)	c-c domains	Major head protein
6	1	5365-5655	11.0 (96)	Gp7 ϕ3626 *C. perfringens* (2e-3) Phage QLRG family (1e-7)		
7	3	5652-5975	12.1 (107)	BA_4559 *B. anthracis* (6e-5) Bacteriophage head-tail adaptor (4e-9)	Pyrophosphatase domain	Putative head-tail adaptor
8	1	5968-6408	16.2 (146)	BA_4558 *B. anthracis* (2e-16) CRPp0301 (2e-5)	s	Uncharacterized protein
9	3	6405-6764	13.9 (119)	BA_4557 *B. anthracis* (4e-15) CRPp0346 (5e-8)		Uncharacterized protein
10	3	6765-7373	22.9 (202)	Chte_p_1640 *C. thermocellum* (9e-29) CRPp0161 (6e-11)	c-c domain	Major tail protein
11	1	7423-7740	11.8 (105)	BA_4555 *B. anthracis* (2e-3)		
12	3	7770-7946	7.0 (58)			
13	1	7963-11814	139.4 (1283)	BA_4552-BA_4554 *B. anthracis* (>6e-25), CRPp0381 (4e-40)	c-c domains	Tail protein
14	3	11829-13319	56.8 (496)	BA_4550 *B. anthracis* (e-153), CRPp0325 (2e-4)	c-c domain, Pyridoxal-phosphate binding domain	Putative tail component protein

Wp	Frame	Position	Size	Matches (E value)*	Structure†	Function‡
15	2	13316-17311	149.8 (1331)	BA_4578-BA_4579 *B. anthracis* (>5e-20), CRPp0329 (2e-36)	c-c domains	Similar to myosin heav chain
16	1	17350-17775	15.0 (141)	BA_4545 *B. anthracis* (9e-57), Phage-related holin (3e-28)	3 TM domains	Holin
17	3	17775-18476	26.3 (233)	BA_4545 *B. anthracis* (1e-112) Cell wall amidase (2e-43)		Lysin
18	-1	19031-18534	18.0 (165)		c-c domain, 1 TM domain	May be a lipoprotein
19	-3	19230-19018	8.3 (70)	BA_4541 *B. anthracis* (4e-13) Helix-turn-helix XRE-family (1e-4)	HTH domain only	
20	1	19414-19722	12.3 (102)	BA_4540 *B.anthracis* (2e-40)		
21	3	19719-19901	6.7 (60)	BA_4539 *B. anthracis* (3e-12)	2 TM domains	
22	3	19911-21200	49.3 (414)	BA_4538 *B. anthracis* (e-168) FtsK/SpoIIIE family: C- term. (1e-15)	c-c domain, P-loop (ATP/ GTP binding)	DNA translocation? Integration?
23	1	21178-21759	25.2 (175)	BC1920 *B. cereus* (2e-61)	c-c domain	
24	-3	22029-21793	8.6 (78)	BC1914 *B. cereus* (5e-18)		
25	2	21863-22090	8.4 (75)		2 TM domains	
26	2	22325-23188	33.3 (287)	BC4930 *B. cereus* (6e-11)	c-c domain	Transcriptional effects?
27	2	23264-24709	56.4 (481)	Chte1631*C. thermocellum* (2e-30) PinR_Site-specific recombinases (4e-20)	c-c domains	Integrase
28	2	24812-26146	51.2 (444)	Orf4 *B. thuringiensis* pAW63 plasmid (3e-6)		Absent from γ
28.1	-3	26488-26844	13.6 (115)	BC2558 *B. cereus* (1e-11) Helix-turn-helix Cro and CI family (2e-7)	c-c domain	CI-like DNA binding role? Absent from γ
29	1	27004-27231	8.8 (75)	BC2559 *B. cereus* (7e-5) Helix-turn-helix Cro and CI family (0.008)	HTH domain	Cro-like DNA binding role?
30	1	27244-27429	7.2 (61)	BA_4542 *B. antrhacis* (7e-5)		
31	2	27674-28489	31.0 (271)	Orf6 *L. lactis* ϕbIL285 (2e-12), CRPp0355 (6e-17)	c-c domain	Anti-repressor
32	2	28556-29209	25.7 (217)	Orf16 *L. lactis* ϕbIL312 (4e-12)		
33	1	29338-30285	37.1 (315)	DnaA phage analogs Orf11 *L. lactis* ϕrlt (1e-18)	c-c domain AT-rich repeats	Phage replication; Origin of replication
34	1	30301-31212	34.9 (303)	DnaC DNA replication protein (3e-10) Ntp *Lactobacillus* ϕg1e (6e-29)	c-c domain; P-loop (ATP-GTP binding)	Phage replication
35	1	31231-31464	9.2 (77)			
36	2	31457-32203	28.1 (248)	BA_4585 *B, anthracis* (6e-41) FliA family of sigma factors (1e-20)		Transcriptional effects
37	1	32200-32676	19.0 (158)			
38	3	32736-33278	21.1 (180)	BA_5241 *B. anthracis* (2e-20)		
39	1	33514-34446	28.7 (310)	Bcol14-2 *B. thuringiensis* pTX14-2 plasmid (6e-83)	4 collagen-like triple helix repeats	Spore surface antigen
40	3	34440-34931	16.8 (163)	BC4769 *B. cereus* (2e-26) C.-term half of collagen triple helix repeat protein	4 TM domains	
41	-2	35903-35379	21.2 (191)	CTC01899 *C.tetani* (2e-54) Mannose-6-phosphate isomerase (5e-19)		Nutrient acquisition or a role in surface carbohydrate structure
42	1	36490-36660	6.5 (56)			
43	3	36849-37154	11.9 (101)		c-c domain	
44	1	37147-37386	9.3 (79)			
45	2	37823-38230	16.0 (135)	BC3700 *B. cereus* (1e-39)	c-c domains	
46	1	38350-38574	8.5 (74)		c-c domains	
47	1	38581-38802	8.2			30% identity to C-term.

Wp	Frame	Position	Size	Matches (E value)*	Structure†	Function‡
			(73)			half of hamster IL-6
48	2	38807-39211	15.6 (134)			
49	1	39316-39507	7.2 (63)			
50	2	39527-39781	10.2 (84)		c-c domains	
51	2	39980-40171	7.3 (63)			
52	3	40152-40439	10.6 (95)	BA_4569 *B. anthracis* (5e-4)	c-c domains	
53	2	40439-40822	15.5 (128)	Gp50 ϕ3626 *C. perfringens* (6e-13) McrA restriction endonuclease (2e-5)		endonuclease

*No entry indicates no significant homologies based on a protein-protein BLAST search. First line indicates the protein name, organism of origin, and BLAST E value for most significant hit. Second line indicates, if detected, the pfam conserved domain and E value or the cluster of related viral proteins (CRP) designation and E value.
†Indicates a significant protein structure or motif detected by bioinformatics analysis. Abbreviations are used: c-c domain/s, indicating the significant likelihood of one or more coiled-coil domains; TM, for transmembrane domain; and HTH, for helix-turn-helix.
‡Indicates putative function based on homologies detected with proteins of known function.

POLYNUCLEOTIDE SEQUENCE COMPARISON

[77] Similar Features of γ and W Bacteriophage Sequences

[78] The genomic sequences of both γ and W were determined and shown by pairwise comparison to be 100% identical with exceptions at four loci (**Fig. 5A**). The G_C contents of γ and W were 35.1% and 35.3%, respectively, similar to that of the *B. anthracis* genome (36.4%). Complementary 9 bp 5'-single-stranded cohesive ends (cos sites) flanked both phage. The γ phage encoded 53 ORFs over 37,367 bp, while the parental W phage had 54 ORFs within its 40,864 bp genome.

[79] A common feature of the lambdoid genomes is a genetic mosaicism that results from rampant recombination and the horizontal transfer of functional gene modules (discrete transcriptional units containing one or more genes) among related phage genomes infecting, perhaps, a disparate range of bacterial organisms. As such, the genomes appear as a 'pasting' of modules from different sources, encoding part or all of each of the basic phage functions, including capsid building, host lysis, lysogeny, and replication. The architecture of the γ and W genomes is consistent with this model. The virion structural and host lysis proteins of γ and W (ORFs 1-17), are the most well conserved components, similar in both sequence and gene order to phage elements encoded by phages ϕ3626 of *Clostridium perfringens*, ϕ105 of *B. subtilis* or ϕ4537 and ϕ4241 prophages deduced from the *B. anthracis* genome (**FIG. 5B**). The lysogeny genes

(ORFs 26-30) are divergent, showing homology for phage elements of *B. cereus*, C. thermocellum, Lactococcus lactis *and a plasmid gene of* Bacillus *spp.* The replication module (ORFs 31-34) is primarily similar to replication elements from phage of *L. lactis* and lactobacilli. In this manner, the functional genes of γ and W are indeed an assembled mosaic. A high proportion of genes (61 and 62% for the W and γ genomes, respectively) are similar to phage proteins from $Gram^+$ spore forming bacteria. Twenty-one of these genes are similar to elements of *B. anthracis* ϕ4537, and within this group, eight genes are found nowhere else. Alignment of the W phage and with the complete ϕ4537 genome and the late genes of ϕ4241 are presented (**FIG. 5C**) to illustrate the extent of this homology and the likelihood that W (and thus γ) arose from a common precursor of these phage. This divergence was likely not recent owing to the notable difference between the W and ϕ4537 genomes seen in **FIG. 5C**, and the presence of twelve largely unlinked γ and W loci are novel genes unrelated to known phage and host proteins. One feature of the γ and W genomes is the presence of 8 loci between the Orf17 amidase and the lysogenic module (starting at Orf26), which are similar only to B. antracis and *B. cereus* phage. Notably in *Streptococcal* phage, this position often encodes genes not for phage function, but for lysogenic conversion of the host. This region in γ and W notably encodes two host membrane proteins and a 1242 bp gene homologous to the 5' half of the bacterial host cell division protein FtsK. The presence of an FtsK homolog in *B. anthracis* and transition strains may relate to the notable chain-like morphology of these organisms. Downstream of the replication module is another notable B. antracis phage-specific host factor, Orf35, encoding a homolog of the sporulation sigma factor sigma F. Sigma F directs the RNA polymerase holoenzyme to a specific set of gene promoters within the developing spore of Bacillus spp. The presence of such a regulatory factor in W phage suggest that lysogeny may be accompanied by alterations in host gene expression.

[80] Features of which differ between γ and W Bacteriophage Sequences

[81] Differences between γ and W, were observed with respect to the phage and to the host. Four changes have occurred in γ (compared to W) in the 50 years since its isolation and use as a diagnostic phage for *B. anthracis*. Without being bound by theory, it is believed that two alterations in the lysogeny module relate to the conversion of γ from a lysogenic to a lytic phage. It is further believed that a set of alterations with a single tail fiber gene explains the reported alteration in host specificity (ability of γ to infect encapsulated *B. anthracis*) and defines the gene which is essentially the basis for the widespread use of γ as a diagnostic tool. The last alternation is believed to be particularly significant, and involves the replacement of a three gene island in W with an alternate three gene island in γ.

[82] *Changes in the lysogeny module*

[83] Changes in the gamma lysogeny module (ORFs 26-32) may explain the derivation of gamma from W. The lysogeny region is a known hotspot for recombination in several phage, including W. The decision between lytic and lysogenic growth is often influenced by a genetic switch region encoding two divergently transcribed small DNA binding repressor proteins, which represent functional homologs of the well studied CI and Cro proteins of L phage. In phage W, the CI and Cro-like functions are likely encoded by wp28.1 and wp29, with Wp28.1 (CI-like) required for repressing the lytic proliferation genes and promoting lysogeny and Wp29 (Cro-like) required for repressing expression of the lysogeny module and promotin lytic growth. In the gamma phage, the lytic-only variant of W, both Wp28.1 and the adjacent gene Wp28 have been lost as part of a 2003 bp deletion that fused the 5' third of Wp28.1 to a short peptide sequence between Wp27 and Wp28, creating Gp28, a presumed gene fragment encoding only a partial heli-turn-helix DNA binding motf. In addition to this, there is a 28 bp deletion in an intergenic region between ORFs25 and 26, which is immediately adjacent to the phage attachment (att) site, which is required for insertion of the phage into the host genome during the establishment of the lysogenic state. Without being limited to theory, it is believed that the gamma bacteriophage has

developed as a lytic variant through two separate deletion events at sites required for lysogenic functions.

[84] *Changes in the Orf14 tail fiber gene*

[85] We sought to identify γ encoded genes that specify the interaction with the surface of *B. anthracis*. Elements which are the basis of diagnostic tools and also key to the phage infection cycle. Two likely candidates observed in genome — PlyG (lysine known to bind Ba, however it has same sequence in both γ and W) and Orf14 (putative tail fiber, which has undergone major change in gamma compared to W). We investigated the ability of each to bind Ba and RSVF1 using GFP protein fusions.

[86] At least 69 missense mutations have occurred in γ ORF14 (referred to as Gp14) since its isolation from W 50 years ago. The resultant proteins differ by 24 amino acid residues (92% identity), likely affecting structural changes in the binding domain need for improved infection of a *B. anthracis* host. The gene appears to have arisen specifically in the Ba phage through insertion of a novel binding module into a tail fiber found in many bacillus phage.

[87] *Three gene island in W phage polynucleotide sequence*

[88] The W phage as a 2824 bp three gene island (ORFs 39-41) encoding a putative spore surface antigen, a transmembrane domain that may be expressed with the surface antigen as part of a translational frameshift mechanism, and an enzyme (often associated with pathogenicity islands) which is a mannose-6-phosphate isomerse. The spore antigen appears to be similar, but not identical to, fibrous appendages that are found on the surface of spores, are the dominant surface antigen of spores, and are likely involved in the initial infection process of *Bacillus anthracis*. The mannose-6-phosphate isomerase is often considered a horizontally transferred virulence associated gene involved in generating alterations in surface carbohydrate structure in Gram bacteria. This three gene island appears to encode proteins not required directly for the phage lifecycle, but are rather of use to the host (lysogenic conversion genes). The gamma phage has lost this island probably due to recombination with a three gene segment in *B. anthracis* ϕ4567. This 1360 bp segment (replacing the 2824 bp W island) is

99% identical to sequence in γ. This island encodes two proteins found only in Ba phage, and also a Fosfomycin resistance gene. The Fos gene (Gp41) is similar to this family of proteins, which act as glutathione S transferases. Similar genes are found in most bacteria, however, only in Ba is it phage encoded. Most soil bacteria examined (*Clostridium spp.*, *Bacillus subtilis*, *Bacillus cereus*, have about 10-15 glutathione S transferase-like genes, while *B. anthracis* has almost 40.

[89] Other Polynucleotide Sequences

[90] The invention also relates to nucleotide sequences which have a substantial identity with the nucleotide sequences described herein; particularly preferred are nucleotide sequences which have at least about 10%, preferably at least about 20%, more preferably at least about 30%, more preferably at least about 40%, even more preferably at least about 50%, yet more preferably at least about 70%, still more preferably at least about 80%, and even more preferably at least about 90% identity, or 95% identity or more, with nucleotide sequences described herein. Particularly preferred in this instance are nucleotide sequences encoding polypeptides having an activity of a polypeptide described herein. For example, in one embodiment, the nucleotide sequence encodes a DNA polymerase, 3'-5' exonuclease, 5'-3' exonuclease (RNase H), DNA helicase, or RNA ligase, as described below. In a preferred embodiment, the nucleotide encodes a DNA polymerase lacking exonuclease domains, or a 3'-5' exonuclease lacking DNA polymerase domain, as described below.

[91] To determine the percent identity of two nucleotide sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps may be introduced in the sequence of a first nucleotide sequence). The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity=# of identical positions/total # of positions.times.100).

[92] The determination of percent identity between two sequences may be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin et al., Proc. Natl. Acad. Sci. USA, 90:5873-5877 (1993). Such an algorithm is incorporated into the NBLAST program which may be used to identify sequences having the desired identity to nucleotide sequences of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST may be utilized as described in Altschul et al., Nucleic Acids Res, 25:3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., NBLAST) may be used. See the programs provided by National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health. In one embodiment, parameters for sequence comparison may be set at W=12. Parameters may also be varied (e.g., W=5 or W=20). The value "W" determines how many continuous nucleotides are preferably identical for the program to identify two sequences as containing regions of identity.

[93] One skilled in the art will recognize that the DNA mutagenesis techniques described here may produce a wide variety of DNA molecules that code for a bacteriophage lysin specific for *Bacillus anthracis* yet that maintain the essential characteristics of the lytic protein. Newly derived proteins may also be selected in order to obtain variations on the characteristic of the lytic protein, as will be more fully described below. Such derivatives include those with variations in amino acid sequence including minor deletions, additions and substitutions.

[94] While the site for introducing an amino acid sequence variation is predetermined, the mutation per se does not need to be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed protein variants screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence as described above are well known.

[95] Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. Deletions or insertions may be in single form, but preferably are made in adjacent pairs, i.e., a deletion of 2 residues or insertion of 2 residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a final construct. Obviously, the mutations that are made in the DNA encoding the protein preferably does not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure (EP 75,444A).

[96] Substitutional variants are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions may be made in accordance with the following Table 3 when it is desired to finely modulate the characteristics of the protein. Table 3 shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions.

Table 3

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln, his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	pro
His	asn; gln
Ile	leu, val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

[97] Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than in Table 3, i.e., selecting residues that differ more significantly in their effect on maintaining: (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation; (b) the charge or hydrophobicity of the molecule at the target site; or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which: (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histadyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

[98] The effects of these amino acid substitutions or deletions or additions may be assessed for derivatives of the lytic protein by analyzing the ability of the derivative proteins to complement the sensitivity to DNA cross-linking agents exhibited by phages in infected bacteria hosts. These assays may be performed by transfecting DNA molecules encoding the derivative proteins into the bacteria as described above.

[99] Having herein provided nucleotide sequences that code for lytic enzyme genetically coded for by a bacteriophage specific for *Bacillus anthracis* and fragments of that enzyme, correspondingly provided are the complementary DNA strands of the cDNA molecule and DNA molecules which hybridize under stringent conditions to the lytic enzyme cDNA molecule or its complementary strand. Such hybridizing molecules include DNA molecules differing only by minor sequence changes, including nucleotide substitutions, deletions and additions. Also contemplated by this disclosure are isolated oligonucleotides comprising at least a segment of the cDNA molecule or its complementary strand, such as oligonucleotides which may be employed as effective DNA hybridization

probes or primers useful in the polymerase chain reaction. Hybridizing DNA molecules and variants on the lytic enzyme cDNA may readily be created by standard molecular biology techniques.

[100] The detection of specific DNA mutations may be achieved by methods such as hybridization using specific oligonucleotides (Wallace et al. (1986). Cold Spring Harbor Symp. Quant. Biol. 51:257-261), direct DNA sequencing (Church and Gilbert (1988). Proc. Natl. Acad. Sci. USA 81:1991-1995), the use of restriction enzymes (Flavell et al. (1978). Cell 15:25), discrimination on the basis of electrophoretic mobility in gels with denaturing reagent (Myers and Maniatis (1986). Cold Spring Harbor Symp. Quant. Biol. 51:275-284), RNase protection (Myers et al. (1985). Science 230:1242), chemical cleavage (Cotton et al. (1985). Proc. Natl. Acad. Sci. USA 85:4397-4401) (incorporated herein by reference), and the ligase-mediated detection procedure (Landegren et al., 1988).

EXPRESSION VECTORS

[101] The invention also relates to expression vectors containing a nucleic acid sequence encoding a polypeptide described herein (or an active derivative or fragment thereof), operably linked to at least one regulatory sequence. Many expression vectors are commercially available, and other suitable vectors may be readily prepared by the skilled artisan. "Operably linked" is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleic acid sequence. Regulatory sequences are art-recognized and are selected to produce the polypeptide or active derivative or fragment thereof. The term "regulatory sequence" includes promoters, enhancers, and other expression control elements which are described in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). For example, the native regulatory sequences or regulatory sequences native to bacteriophages γ and W may be employed. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of polypeptide desired to be expressed. For instance, the polypeptides of the present invention may be

produced by ligating the cloned gene, or a portion thereof, into a vector suitable for expression in an appropriate host cell (see, for example, Broach, et al., Experimental Manipulation of Gene Expression, ed. M. Inouye (Academic Press, 1983) p. 83; Molecular Cloning: A Laboratory Manual, 2nd Ed., ed. Sambrook et al. (Cold Spring Harbor Laboratory Press, 1989) Chapters 16 and 17). Typically, expression constructs will contain one or more selectable markers, including, but not limited to, the gene that encodes dihydrofolate reductase and the genes that confer resistance to neomycin, tetracycline, ampicillin, chloramphenicol, kanamycin and streptomycin resistance. Thus, prokaryotic and eukaryotic host cells transformed by the described expression vectors are also provided by this invention. The host cells may be transformed by the described vectors by various methods (e.g., electroporation, transfection using calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection, infection where the vector is an infectious agent such as a retroviral genome, and other methods), depending on the type of cellular host. The nucleic acid molecules of the present invention may be produced, for example, by replication in such a host cell, as described above. Alternatively, the nucleic acid molecules may also be produced by chemical synthesis.

PROBES

[102] The isolated nucleic acid molecules and vectors of the invention are useful in the manufacture of the encoded polypeptide, as probes for isolating homologous sequences (e.g., from other bacteriophage species), as well as for detecting the presence of the bacteriophage in a culture of host cells.

[103] The nucleotide sequences of the nucleic acid molecules described herein (e.g., a nucleic acid molecule comprising any of the open reading frames shown in **Table 1** or **Table 2** may be amplified by methods known in the art. For example, this may be accomplished by e.g., PCR. See generally PCR Technology: Principles and Applications for DNA Amplification (ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992); PCR Protocols: A Guide to Methods and Applications (eds. Innis, et al., Academic Press, San Diego, Calif., 1990); Mattila et al., Nucleic Acids Res. 19, 4967 (1991); Eckert et al., PCR Methods and Applications 1, 17

(1991); PCR (eds. McPherson et al., IRL Press, Oxford); and U.S. Pat. No. 4,683,202.

[104] Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, Genomics 4, 560 (1989), Landegren et al., Science 241, 1077 (1988), transcription amplification (Kwoh et al., Proc. Natl. Acad. Sci. USA 86, 1173 (1989)), and self-sustained sequence replication (Guatelli et al., Proc. Nat. Acad. Sci. USA, 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

[105] The amplified DNA may be radiolabelled and used as a probe for screening a library or other suitable vector to identify homologous nucleotide sequences. Corresponding clones may be isolated, DNA may be obtained following in vivo excision, and the cloned insert may be sequenced in either or both orientations by art recognized methods, to identify the correct reading frame encoding a protein of the appropriate molecular weight. For example, the direct analysis of the nucleotide sequence of homologous nucleic acid molecules of the present invention may be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook et al., Molecular Cloning, A Laboratory Manual (2nd Ed., CSHP, New York 1989); Zyskind et al., Recombinant DNA Laboratory Manual, (Acad. Press, 1988)). Using these or similar methods, the protein(s) and the DNA encoding the protein may be isolated, sequenced and further characterized.

POLYPEPTIDES

[106] The invention additionally relates to isolated polypeptides obtainable from the bacteriophages γ and W. The term, "polypeptide," as used herein, includes proteins, enzymes, peptides, and gene products encoded by nucleic acids described herein. In one embodiment, the invention relates to the polypeptides

encoded by the ORFs as described above in **Table 1** and **Table 2**. The invention relates to polypeptide sequences for the γ-phage and other polypeptides that may hybridize to the polypeptide sequences of the invention, including those of **FIG. 2B**. The invention further relates to polypeptide sequences for the W-phage, such as those in **FIG. 4B**, and other polypeptides that may hybridize to these sequences. Also provided in the present invention are polypeptide sequences for each ORF in **Table 1** and **Table 2**. The invention relates to polypeptides encoding Gp 14 (ORF 14 of g-phage), Wp14 (ORF 14 of W-phage), and Wp38 (ORF 38 of W-phage). Further provided are polynucleotide sequences that hybridize to polypeptide sequences of **FIG. 2B** and **FIG. 4B**.

[107] Also included in the invention are polypeptides which are at least about 60, 70, 80, 90, and 95% identical (i.e., polypeptides which have substantial sequence identity) to the polypeptides described herein. However, polypeptides exhibiting lower levels of identity are also useful, particular if they exhibit high, e.g., at least about 90%, identity over one or more particular domains of the polypeptide. For example, polypeptides sharing high degrees of identity over domains necessary for particular activities, such as binding or enzymatic activity, are included herein. Thus, polypeptides which are at least about 10%, preferably at least about 20%, more preferably at least about 30%, more preferably at least about 40%, even more preferably at least about 50%, yet more preferably at least about 60%, still more preferably at least about 70%, yet more preferably at least about 80%, still more preferably at least about 90%, yet more preferably at least about 95%, still more preferably at least about 80% and even more preferably at least about 97% identity to the polypeptides of the invention, including SEQ ID NO:3 through SEQ ID NO:109, are encompassed by the invention.

[108] Polypeptides described herein may be isolated from naturally-occurring sources (e.g., isolated from host cells infected with bacteriophages γ and W). Alternatively, the polypeptides may be chemically synthesized or recombinantly produced. For example, PCR primers may be designed to amplify the ORFs from the start codon to stop codon. The primers may contain suitable restriction sites for

an efficient cloning into a suitable expression vector. The PCR product may be digested with the appropriate restriction enzyme and ligated between the corresponding restriction sites in the vector (the same restriction sites, or restriction sites producing the same cohesive ends or blunt end restriction sites).

[109] Polypeptides of the present invention may be used as a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using art-recognized methods. They are particularly useful for molecular weight markers for analysis of proteins from thermophilic organisms, as they will behave similarly (e.g., they will not denature as proteins from mesophilic organisms would).

[110] The polypeptides of the present invention may be isolated or purified (e.g., to homogeneity) from cell culture (e.g., from culture of host cells infected with bacteriophages γ and W) by a variety of processes. These include, but are not limited to, anion or cation exchange chromatography, ethanol precipitation, affinity chromatography and high performance liquid chromatography (HPLC). The particular method used will depend upon the properties of the polypeptide; appropriate methods will be readily apparent to those skilled in the art. For example, with respect to protein or polypeptide identification, bands identified by gel analysis may be isolated and purified by HPLC, and the resulting purified protein may be sequenced. Alternatively, the purified protein may be enzymatically digested by methods known in the art to produce polypeptide fragments which may be sequenced. The sequencing may be performed, for example, by the methods of Wilm et al. (Nature 379(6564):466-469 (1996)). The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, i.e., 80, 95 or 99% free of cell component contaminants, as described in Jacoby, Methods in Enzymology Volume 104, Academic Press, New York (1984); Scopes, Protein Purification, Principles and Practice, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), Guide to Protein Purification, Methods in Enzymology, Vol. 182 (1990).

[111] For example, representative proteins expected to be encoded by genes of bacteriophages γ and W include the following: DNA topoisomerase; exonuclease (e.g., 3'-5' exonuclease, 5'-3' exonuclease (RNase H)); helicase; enzymes related to DNA or RNA synthesis (e.g., dCTPase, dUTPase, dCDPase, dUDPase, GTPase, dGTPase, ATPase, dATPase); transposase; reverse transcriptase; polymerase (e.g., DNA polymerase, RNA polymerase); DNA polymerase accessory protein; DNA packaging protein; DNA topoisomerase; RNA polymerase binding protein; RNA polymerase sigma factor; site-specific RNase inhibitor of protease; recombinant protein; alpha-glucosyltransferase; mobility nuclease; endonuclease (e.g., endonuclease II, endonuclease V, endonuclease VII); inhibitor of Lon protease; thymidine kinase; site-specific RNase; N-glycosidase; endolysin; lysozyme; dNMP kinase; DNA ligase; deoxyribonucleotide-3'-phosphatase; ssDNA binding protein; dsDNA binding protein; and RNA ligase.

[112] The polypeptides of the invention may be partially or substantially purified (e.g., purified to homogeneity), and/or are substantially free of other polypeptides. According to the invention, the amino acid sequence of the polypeptide may be that of the naturally-occurring polypeptide or may comprise alterations therein. Polypeptides comprising alterations are referred to herein as "derivatives" of the native polypeptide. Such alterations include conservative or non-conservative amino acid substitutions, additions and deletions of one or more amino acids; however, such alterations should preserve at least one activity of the polypeptide, i.e., the altered or mutant polypeptide should be an active derivative of the naturally-occurring polypeptide. For example, the mutation(s) may preferably preserve the three dimensional configuration of the binding site of the native polypeptide, or may preferably preserve the activity of the polypeptide (e.g., if the polypeptide is a DNA polymerase, any mutations preferably preserve the ability of the enzyme to catalyze combination of nucleotide triphosphates to form a nucleic acid strand complementary to a nucleic acid template strand). The presence or absence of activity or activities of the polypeptide may be determined

by various standard functional assays including, but not limited to, assays for binding activity or enzymatic activity.

[113] Additionally included in the invention are active fragments of the polypeptides described herein, as well as fragments of the active derivatives described above. An "active fragment," as referred to herein, is a portion of polypeptide (or a portion of an active derivative) that retains the polypeptide's activity, as described above.

[114] Homologous proteins and nucleic acids may be prepared that share functionality with such small proteins and/or nucleic acids (or protein and/or nucleic acid regions of larger molecules) as will be appreciated by a skilled artisan. Such small molecules and short regions of larger molecules, that may be homologous specifically are intended as embodiments. Preferably the homology of such valuable regions is at least 50%, 65%, 75%, 85%, and more preferably at least 90%, 95%, 97%, 98%, or at least 99% compared to the polypeptides encoded by a polypeptide sequence selected from SEQ ID NO:3- SEQ ID NO:109. These percent homology values do not include alterations due to conservative amino acid substitutions.

[115] Of course, an epitope as described herein may be used to generate an antibody and also may be used to detect binding to molecules that recognize the lysin protein. Another embodiment is a molecule such as an antibody or other specific binder that may be created through use of an epitope such as by regular immunization or by a phase display approach where an epitope may be used to screen a library if potential binders. Such molecules recognize one or more epitopes of lysin protein or a nucleic acid that encodes lysin protein. An antibody that recognizes an epitope may be a monoclonal antibody, a humanized antibody, or a portion of an antibody protein. Desirably the molecule that recognizes an epitope has a specific binding for that epitope which is at least 10 times as strong as the molecule has for serum albumin. Specific binding may be measured as affinity (Km). More desirably the specific binding is at least 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, or even higher than that for serum albumin under the same conditions.

[116] In a desirable embodiment the antibody or antibody fragment is in a form useful for detecting the presence of the lysin protein. A variety of forms and methods for their synthesis are known as will be appreciated by a skilled artisan. The antibody may be conjugated (covalently complexed) with a reporter molecule or atom such as a fluor, an enzyme that creates an optical signal, a chemilumiphore, a microparticle, or a radioactive atom. The antibody or antibody fragment may be synthesized in vivo, after immunization of an animal, for example, The antibody or antibody fragment may be synthesized via cell culture after genetic recombination The antibody or antibody fragment may be prepared by a combination of cell synthesis and chemical modification.

[117] Biologically active portions of a protein or peptide fragment of the embodiments, as described herein, include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the phage protein of the disclosure, which include fewer amino acids than the full length protein of the phage protein and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein or protein fragment of the disclosure may be a polypeptide which is, for example, 10, 25, 50, 100 less or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, or added may be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the embodiments.

[118] Appropriate amino acid alterations may be made on the basis of several criteria, including hydrophobicity, basic or acidic character, charge, polarity, size, the presence or absence of a functional group (e.g., --SH or a glycosylation site), and aromatic character. Assignment of various amino acids to similar groups based on the properties above will be readily apparent to the skilled artisan; further appropriate amino acid changes may also be found in Bowie et al. (Science 247:1306-1310(1990)). For example, conservative amino acid replacements may be those that take place within a family of amino acids that are related in their side

chains. Genetically encoded amino acids are generally divided into four families: (1) acidic=aspartate, glutamate; (2) basic=lysine, arginine, histidine; (3) nonpolar=alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar=glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine. Phenylalanine, tryptophan and tyrosine are sometimes classified jointly as aromatic amino acids. For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine or a similar conservative replacement of an amino acid with a structurally related amino acid will not have a major effect on activity or functionality.

FUSION PROTEINS

[119] The polypeptides of the invention may also be fusion polypeptides comprising all or a portion (e.g., an active fragment) of the native bacteriophages γ and W polypeptide amino acid sequence fused to an additional component, with optional linker sequences. Additional components, such as radioisotopes and antigenic tags, may be selected to assist in the isolation or purification of the polypeptide or to extend the half life of the polypeptide; for example, a hexahistidine tag would permit ready purification by nickel chromatography. The fusion protein may contain, e.g., a glutathione-S-transferase (GST), thioredoxin (TRX) or maltose binding protein (MBP) component to facilitate purification; kits for expression and purification of such fusion proteins are commercially available. One example of a useful fusion protein is a GST fusion protein in which the polypeptide of the disclosure is fused to the C-terminus of a GST sequence. The polypeptides of the invention may also be tagged with an epitope and subsequently purified using antibody specific to the epitope using art recognized methods. Additionally, all or a portion of the polypeptide may be fused to carrier molecules, such as immunoglobulins, for many purposes, including increasing the valency of protein binding sites. For example, the polypeptide or a portion thereof may be linked to the Fc portion of an immunoglobulin; for

example, such a fusion could be to the Fc portion of an IgG molecule to create a bivalent form of the protein.

[120] Additionally, the nucleotide sequence(s) may be fused to a marker sequence, for example, a sequence which encodes a polypeptide to assist in isolation or purification of the polypeptide. Representative sequences include, but are not limited to, those that encode a glutathione-S-transferase (GST) fusion protein. In one embodiment, the nucleotide sequence contains a single ORF in its entirety (e.g., encoding a polypeptide, as described below); or contains a nucleotide sequence encoding an active derivative or active fragment of the polypeptide; or encodes a polypeptide which has substantial sequence identity to the polypeptides described herein. In a preferred embodiment, the nucleic acid encodes a polymerase (e.g., DNA polymerase); DNA polymerase accessory protein; dsDNA binding protein; deoxyriboncleotide-3-phosphatase; DNA topoisomerase; DNA helicase; an exonuclease (e.g., 3'-5' exonuclease, 5'-3' exonuclease (RNase H)); RNA ligase; site-specific RNase inhibitor of protease; endonuclease; exonuclease; mobility nuclease; reverse transcriptase; single-stranded binding protein; endolysin; lysozyme; helicase; alpha-glucosyltransferase; or thymidine kinase, as described herein. In a particularly preferred embodiment, the nucleic acid encodes a DNA polymerase, 3'-5' exonuclease, 5'-3 exonuclease (RNase H), DNA helicase or RNA ligase. In another particularly preferred embodiment, the nucleic acid encodes a DNA polymerase that lacks exonuclease domains, or a 3'-5' exonuclease that lacks DNA polymerase domain, as described below.

[121] Another embodiment discloses an immunoglobulin fusion protein in which all or part of a polypeptide of the disclosure is fused to sequences derived from a member of the immunoglobulin protein family. An immunoglobulin fusion protein may be incorporated into a pharmaceutical composition and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction in vivo. The immunoglobulin fusion protein may alter bioavailability of a cognate ligand of a polypeptide of the disclosure.

Inhibition of ligand/receptor interaction may be useful therapeutically, both for treating bacterial-associated diseases and disorders for modulating (i.e. promoting or inhibiting) cell survival. Moreover, an immunoglobulin fusion protein of the disclosure may be used as an immunogen to produce antibodies directed against a polypeptide of the disclosure in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands. Chimeric and fusion proteins and peptides of the disclosure may be produced by standard recombinant DNA techniques.

[122] The acts of methods of the present invention may be done in any order, and may have other intervening steps or acts unless otherwise indicated.

[123] It is intended that the foregoing detailed description be regarded as illustrative rather than limiting. The following claims, including all equivalents, that are intended to define the spirit and scope of this invention. Therefore, the embodiment of any figure and features thereof may be combined with the embodiments depicted in other figures. Other features known in the art and not inconsistent with the structure and function of the present invention may be added to the embodiments.

[124] The recitations of "embodiments," "one embodiment," "some embodiments," "other embodiments," "illustrative embodiments," "selected embodiments," "certain embodiments," and "another embodiment" herein are synonymous. All of these recitations refer to illustrative embodiments and are not exclusive of each other or of other embodiments not recited herein. The invention also relates to embodiments that comprise combinations of one or more of the illustrative embodiments described above.

[125] All references cited herein are hereby incorporated into this disclosure in their entirety.

[126] The following Examples are offered for the purpose of illustrating the present invention and are not to be construed to limit the scope of this invention. The teachings of all references cited are hereby incorporated herein by reference in their entirety.

EXAMPLES

[127] *Example 1: Isolation of polynucleotides of γ and W bacteriophages*

[128] The *B. anthracis* strain used in this study lacks the pXO1 and pXO2 virulence plasmids (*Δsterne*), while the *B. cereus* strain used, RSVF1, is an American Type Culture Collection (ATCC) reference strain that is largely genetically indistinct from *B. anthracis* and transition state *B. cereus*. Initially, we determined whether RSVF1 represents a transition strain (or derivative thereof, as it is a laboratory passaged strain) by looking for *B. anthracis* features either shared by or lacking in the transition strains. Like *B. anthracis*, RSVF1 was non-motile, sensitive to the γ lysine, grew in chains, was virulent in mice, and encoded the *csa* operon (products of which modify *B. anthracis* surface carbohydrate), the *Ba813* locus (diagnostic marker for *B. anthracis* and transition state *B. cereus*), and a specific 12 base pair tandem repeat array within *vrrA* (characteristic of certain *B. anthracis* isolates). Unlike *B. anthracis*, but like transition *B. cereus*, RSVF1 lacked pXO1 and pXO2, and had a functional PlcR transcriptional regulator (inactive in *B. anthracis* owing to a single nonsense mutation.). Prophage content differed as well, based on findings that a) PCR analyses of several distinct *B. anthracis* prophage genes yielded no amplification products with RSVF1, and b) long-range repetitive PCR using primers specific for a phage attachment site detects gross genetic polymorphisms between *B. anthracis* and transition state *B. cereus*. RSVF1 does, therefore, represent transition state *B. cereus*, largely distinguished from *B. anthracis* by the absence of plasmid and phage elements.

[129] The γ phage was isolated as variant of W that had a more *B. anthracis*-specific host range (though still infecting transition state *B. cereus*) and, unlike W, infected both encapsulated and unencapsulated bacilli. As such, γ became an important tool for rapid confirmatory clinical diagnosis of *B. anthracis* still in widespread use.

[130] A majority of the γ phage genome (~95%) was sequenced by Genome Therapeutics Corporation (Waltham, MA) using a library of 3.0-3.5kb fragments as templates. This analysis was performed using ABI dye terminator chemistry on

automated MegaBace 1000 (Amersham) machines. Base calls and quality scores were determined using the PHRED program (Ewing and Green, 1998 Genome Res. 8:186-194) and reads were assembled by using PHRAP with default program parameters and quality scores. Closure of numerous gaps and determination of the phage termini were accomplished at The Rockefeller University using a primer walking method and purified γ DNA as template. At The Rockefeller University samples were thermocycled in an ABI GeneAmp PCR System 9600/9700 and the purified extension products were electrophoresed on an ABI Prism 3700 DNA Analyzer. Sequence data was assembled into a completed contig using the SeqMan program (DNASTAR software package). Putative ORFs were determined by both ORF Finder (www.ncbi/nlm.nih.gov) and GeneMark approach of gene prediction (http://opal.biology.gatech.edu/GeneMark/gmhmm2_prok.cgi). The BLAST algorithms, available through NCBI, were used for similarity searches of putative ORFs.

[131] The W phage genome was sequenced completely at The Rockefeller University using a primer walking method. Primer construction was completely based on the γ phage genomic sequence. Sequence was assembled, annotated and analyzed in the same manner used for the γ genome.

[132] *Example 2: Binding of GFP Fusion Proteins*

[133] GFP-PlyG – binds *Bacillus anthracis* ("Ba") in a whole cell manner (**FIG. 6A**), RSVF1 only a polar positions (**FIG. 6B**). Rare RSVF1 derivatives bind whole cell (**FIG. 6C**).

[134] GFP-Gp14 – whole cell binding with Ba (**FIG. 6D**) and polar with RSVF1 (**FIG. 6E**). Ba lysogenized with W is now polar (**FIG. 6F**). Indicates that phage affect alteration in surface distribution of receptor, likely carbonhydrate. This is a major change in *B. anthracis* phenotype associated with W phage infection. Either directly affects carbohydrate structure (W does encode a gene involved in sugar conversion to mannose, a known component of the Ba carbohydrate) or indirectly via a transcriptional regulatory factor.

[135] GFP-PlyG binding (or presumably Gp14) may be used as a diagnostic tool...shows that when Ba is diluted 10,000 fold in a culture of *B. cereus* 10987, it is still readily identifiable by fluorescence (**FIG. 6G, FIG. 6H**). The speed of this binding (seconds) and difficulty in washing it away (can stand up to >5 washes in PBS) suggests that GFP-PlyG may be used as a part of a Ba diagnostic method.

[136] *Example 3: Effects of Lysogeny with W on RSVF1 and B. anthracis*

[137] *Effect on RSVF1:*

[138] No alteration in % sporulation, hemolysis, extracellular lipase or protease secretion, motility or colony morphology. Did notice two main changes, however. 1) The strain no longer grows as a filamentous form, but as a rod shape (**FIG. 7A, FIG. 7B**). 2) the spore structure is altered. The spore has a more mottled or textured appearance by SEM (compared to parental non-lysogen) (**FIG. 7C, FIG. 7D**). By negative staining TEM, there is no difference in exosporial structure or surface components. There is however, a change in the coat structure. The coat is more osmiophillic and consist on multiple darkened outer layers, with surface knob-like extensions or striations (**FIG. 7E, FIG. 7F**).

[139] *Effect on B. anthracis:*

[140] No alteration in % sporulation, motility, colony morphology or shape. However, the strain is now alpha hemolytic. Infection has activated a cryptic hemolysin. Transcriptional activators in the W phage are likely responsible. A hallmark of Ba is it lack of hemolysis, despite having hemolysin. The lysogen definitely now has hemolytic activity on plates. A quantitative analysis was performed based on a technique described by Mignot et al. 2001 (Mol Microbiol 42:1189) in which the lysogen had 64 units of activity and the parental strain (no phage) had 0 units. The lysogen is also weakly but definitely activated for extracellular protease and lipase activity on plates. These features are all more Bc like. An entire regulon of extracellular activities is encoded in both Ba and Bc, however, they are only expressed in Bc (repressed in Ba) due to a point mutation in a transcriptional regulator PlcR in Ba. What we see here is the weak activation of this Bc-like regulon in Ba. Either a phage transcriptional regulator is doing this, or the phage integration is activating some downstream regulator.

[141] *Example 4 Gp14 ORF used with primers*

[142] The entire *gp14* ORF was PCR amplified with primers flanking the 1.5 kb locus, using purified γ phage genomic DNA as template. The primers used were as follows:

[143] 5' ACAGATATCTTGGGGAAACTTAGTTTTACTT 3' (SEQ ID NO:110)

[144] 5' CCCAAGCTTTCATCTATATCTCTCCCTATAACTGA 3' (SEQ ID NO:111)

[145] The *Eco*RV and *Hin*dIII 'sticky ends" were used to clone the 1.5 kb amplification product at the 3' end of *gfpmut2* (GenBank nucleotide accession number AF302837) in plasmid pBAD24::*gfp* digested with *Sma*I and *Hin*dIII. The reference for pBAD24 is Guzman et al., 1995 J. Bacteriol. 177:4121-4130. The reference for the *gfpmut2* is Cormack et al., 1996 Gene 173:33-38. The cloning described above yields a gfpmut2-gp14 translational fusion. The *gfpmut2-gp14* construct was excised with *Eco*RI-*Hin*dIII and cloned into the *Eco*RI-*Hin*dIII sites of the vector pBAD/His (Invitrogen). This creates an in frame His tagged fusion, which was subsequently purified by affinity chromatography. The purified His-GFP-Gp14 fusion protein was used to label both *B. cereus* 4342 and *B. anthracis* Δsterne. Exponential phase bacteria growing in BHI media were washed with PBS and concentrated 10 fold and fixed in a 3% formalin in PBS solution for 20 minutes at room temperature. A 100 μl aliquot was then incubated with 100 μl of the GFP fusion protein for 5 minutes at 4C. The cells were washed with PBS, mounted in SlowFade (Molecular Probes, Inc.), and examined by fluorescence microscopy.

[146] The His-GFP-PlyG fusion protein was constructed, purified, and analyzed in the exact same manner, with the exception that the plyG ORF was amplified with the following primers:

[147] 5' gaagatatcatgttcagtaatggaaatcca 3' (SEQ ID NO:112)

[148] 5' accaagcttttatttaacttcataccaccaac 3' (SEQ ID NO:113)

[149] *Prohpetic Example 5: Use of Wp38 as a means to deliver antigens to the B. anthracis spore surface.*

[150] This may be used for vaccine delivery of an anthrax antigen delivered to the surface of an anthrax spore resulting in a vaccine that may protect against both vegetative anthrax and its spores. Wp38 is encoded in the W phage and is similar to a family of spore surface proteins encoded within the *B. anthracis* and *B. cereus* genomes. It is likely expressed from a sporulation-specific promoter and is integrated into the spore exosporium facing the extracellular environment. Since it is not essential for spore formation and resistance properties, we may modify the wp38 sequence, through fusions to exogenous proteins, thereby effecting delivery of the exogenous proteins to the spore surface.

[151] *Prophetic Example 6: Use of the W lysogenic phage as a means to deliver novel DNA sequences to the B. anthracis genome, and to express those sequences.*

[152] Since the W phage genome is stably maintained in *B. anthracis*, we may genetically modify the phage (ie, insert genes of interest) and allow the recombinant phage to infect and be maintained within the bacterium. If the inserted gene is cloned downstream of an inducible promoter also engineered into the W phage, then an expression system is established. We may thus express foreign genes of interest within *B. anthracis*. Expression may be induced either during vegetative growth or during sporulation. With the addition of signal peptide-encoding sequences to the foreign gene, their protein products may be directed to the vegetative cell surface, or into the bacterial supernatant.

[153] *Prophetic Example 7: Use of the W or gamma phages as tools for intact phage therapy.*

[154] Even though we are not involved in phage therapy, those interested could use these phage in their application. Highly purified phage stocks may be used either alone, or in combinations with other *B. anthracis*-specific phage to kill and clear *B. anthracis* during systemic anthrax infections. This therapy may be performed alone or in conjunction with antibiotic and/or anti-toxin treatment. The

phage stocks may also be used to kill or clear *B. anthracis* from contaminated environmental surfaces or from production facilities.

[155] *Example 8: Use of the gamma or W phage tail protein (Gp14 and Wp14, respectively) as a tool to detect B. anthracis in environmental or clinical samples as a disgnostic .*

[156] A Gp14 fusion with GFP has been constructed and shown to specifically bind the surface of *B. anthracis* and relate to a detectable fluorescent signal. This binding is rapid, requiring incubation of the fusion protein and bacteria for only 1 minute. This binding is specific, as it may be used to readily detect a fluorescing *B. anthracis* rod among a background of non *B. anthracis* bacilli, where the *B. anthracis* is diluted 1:10,000.

[157] *Example 9: Study of the Specificity of γ-phage for B. anthracis and strength of γ-phage*

[158] The gamma phage was isolated from *Bacillus anthracis* and was obtained from Hans W. Ackermann (Laval University, Quebec, Canada). Phage susceptibilities were initially tested by adding 10 ml of high titer gamma aliquots to fresh lawns of strains indicated in Table 1; clearance after 16 h growth indicated susceptibility. A high titer phage stock containing 2.2 $x10^{10}$ plaque forming units (pfu) ml^{-1} was prepared using RSVF1 by a previously described method (Loeffler, J. M., Nelson, D. & Fischetti, V. A. Rapid killing *of Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. Science 294, 2170-2 (2001)). A pfu is a single phage that forms a small clearing zone, or plaque, after successive rounds of infection, growth, and release on lawns of susceptible bacteria. The RSVF1-derived phage stock was used in titer determinations.

[159] To study the specificity and strength of the gamma phage and the resulting lytic enzyme, different strains of Bacillus were prepared. Most strains were grown at 30°C in Luria broth (LB) or brain-heart infusion broth (BHI), supplemented with 1.5% agar when needed. Analyses involving E. coli XL1-Blue (Stratagene) were performed at 37°C, while *B. stearothermophilis* was handled at 55°C. Strain RSVF1 is a streptomycin resistant derivative of *B. cereus* reference

strain ATCC 4342. Despite the similarities between RSVF1 and *Bacillus anthracis*, important genotypic differences exist, and RSVF1 is not a misidentified *Bacillus anthracis* stain (Pannucci, J., Okinaka, R. T., Sabin, R. & Kuske, C. R. *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. J Bacteriol 184, 134-41 (2002); Helgason, E., Caugant, D. A., Olsen, I. & Kolsto, A. B. Genetic structure of population of *Bacillus cereus* and B. thuringiensis isolates associated with periodontitis and other human infections. J Clin Microbiol 38, 1615-22 (2000); Ticknor, L. O. et al. Fluorescent Amplified Fragment Length Polymorphism Analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* Soil Isolates. Appl Environ Microbiol 67, 4863-73 (2001)). Analysis of the vrrA locus of RSVF1 was performed as described (Jackson, P. J. et al. Characterization of the variable-number tandem repeats in vrrA from different *Bacillus anthracis* isolates. Appl Environ Microbiol 63, 1400-5 (1997).). *Bacillus anthracis* manipulations were provided by Leonard W. Mayer (Centers for Disease Control, Atlanta, Georgia) and Abraham L. Turetsky (Aberdeen Proving Grounds, Aberdeen, Maryland). These bacterial strains were then exposed to gamma phage.

[160] *Example 10: studies with the lysin produced by the γ-phage*

[161] It was found that RSVF1 was sensitive to infection by γ-phage, and exhibited matt colony morphology, a filamentous structure, and repetitive sequences in the hypervariable vrrA locus which are all characteristic of *Bacillus anthracis*. The lytic activity of PlyG (the gamma lysin produced by gamma phage) was examined by exposing a panel of the indicated liquid bacterial cultures to either PlyG (20 units) or phosphate buffer. The fold killing represents the decrease in bacterial viability determined 15 minutes post-lysing compared to the buffer treatment. The "Bc" and "Bt" prefixes indicate strains as either *B. cereus* or *B. thuringiensis*, respectively. RSVF1 has no virulence plasmids, but is nonetheless highly related to *Bacillus anthracis* and a suitable gamma phage host.

[162] A phenotypic screen was used to identify gamma phage proteins that lyse RSVF1 "from without." An induced gamma phage expression library in an

E. coli background was permeabilized and overlaid with a RSVF1 lawn. gamma genomic DNA was isolated using the l Maxi kit of Qiagen Inc. 5 mg aliquots of gamma DNA were partially digested with Tsp509I and cloned fragments ranging from 0.5-3.0 kb into the EcoRI site of the arabinose-inducible expression vector pBAD24. The resulting expression library was then transformed into *E. coli* XL1-Blue and screened for lysin activity on glass LB plates containing 100 mg ml^{-1} ampicillin and 0.25% arabinose. The induced library was permeabilized with chloroform vapors and overlaid with exponential phase RSVF1 in 0.75% LB agar. After a 24 h incubation, distinct clearing, or lytic, zones were identified over library members. Corresponding plasmid DNA was prepared and sequenced. DNA sequence analysis and manipulations required the BLASTP (NCBI), ORF finder (NCBI), and SeqMan 5.0 (Dnastar Inc.) programs.

[163] One of the pBAD24::plyG constructs recovered in the library search and encoding only the plyG ORF was used as a source of recombinant PlyG. Expression was induced with 0.25% L-arabinose in an overnight LB culture supplemented with ampicillin 100 mg ml^{-1}. Cells were washed, resuspended in 50 mM Tris, pH 8.0, and lysed with chloroform added to a concentration of 16.6%. Cellular debris and chloroform were removed by centrifugation, yielding the crude PlyG supernatant. The cationic nature of PlyG enabled it to pass through a HiTrap Q Sepharose XL column (Amersham Biosciences), which bound to most contaminants. The enzyme was further purified by application to a Mono S HR 5/5 column (Amersham Biosciences) and elution in a linear gradient containing 1 M NaCl. Active fractions were pooled and purity was assessed by gel electrophoresis and chromatography on a Superose 12 column (Amersham Biosciences) equilibrated with gel filtration standards (Bio-Rad).

[164] Clones that yielded lytic zones all contained a 702 bp gamma ORF encoding a product homologous to lysins called N-acetylmuramoyl-L-alanine amidases. TP21 and f 105 indicate *B. cereus* and *B. subtilis* phage amidases, respectively. CwlA and ClyA are encoded in the *B. cereus* and *B. subtillis* genomes, respectively. The dark shading represents sequence identity and the lighter shading represents similarity. Homology is restricted to their catalytic

NH2-terminal halves, and absent in the lysin-specific COOH-terminal binding domains. Recombinant gamma lysin (called PlyG, for phage lysin gamma) was purified to homogeneity by column chromatography using Coomassie Blue-stained, SDS-Page of purified PlyG. The molecular mass was estimated based on the positions of Kaleidoscope (Bio-Rad) standards that are not shown. The N-terminal sequence of this band corresponds to the predicted PlyG sequence. Gel filtration confirmed a predicted size of ~27 kDa, and suggests that PlyG acts as a monomer and is not proteolytically processed.

[165] *Example 11: In Vitro Lysin Activity*

[166] Activity was examined in several ways. A Spectramax Plus 384 spectrophotometer (Molecular Devices) was used to follow the drop in OD_{650} of logarithmic phase RSVF1 incubated for 15 min at 37°C with serial dilutions of purified PlyG. Enzyme activity in units ml^{-1} was then determined as described (Nelson, D., Loomis, L. & Fischetti, V. A., Prevention and elimination of upper respiratory colonization of mice by *group A streptococci* by using a bacteriophage lytic enzyme. Proc Natl Acad Sci U S A 98, 4107-12 (2001)). It was estimated that 1 unit of enzyme corresponded to 1 mg of PlyG. A crude measure of lysin specificity was performed in which 10 ml drops of purified PlyG (0.5 units) were applied to fresh lawns derived from the indicated strains. After overnight incubation, the appearance of clearing zones was used to assess activity. A liquid killing assay was also used, in which 1.0 ml of logarithmic phase cells (~1.0×10^8 cells) was treated with the indicated amounts of PlyG for 15 min at 37°C; at the indicated time points, samples were removed, washed to remove lysin, and plated for enumeration. As a measure of PlyG-directed lysis, ATP released from dying cells was indirectly measured in a reaction containing a luciferin/luciferase reagent and a microluminometer (PROFILE-1 reagent kit and model 3550i luminometer, New Horizons Diagnostics Corp.) according to the manufacturers protocol. In brief, vegetative cells of the indicated strains were immobilized on 0.45 uM filters at the base of a 0.4 ml reaction chamber. The immobilized cells were washed twice with somatic cell releasing agent to remove impurities and 0.1 ml of PlyG in phosphate buffer was added for 2 min. 0.05 ml of the lucifierin/luciferase reagent

provided with the kit was added and immediately assayed at room temperature for 10 sec. All samples were tested five times. The relative light units released by RSVF1 were consistently ten to twenty percent of its total releasable light (as determined using a strong detergent mixture provided with the kit).

[167] RSVF1 was as sensitive to PlyG killing as a set of *Bacillus anthracis* isolates from America, Europe, Asia and Africa (13). *B. cereus* 10987, a rare strain closely related to *Bacillus anthracis*, was slightly susceptible to PlyG, while all other strains examined were resistant. Helgason, E. et al. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*--one species on the basis of genetic evidence. Appl Environ Microbiol 66, 2627-30 (2000). A more sensitive test of PlyG-mediated killing was evaluated in buffer containing ~5.0×10^7 bacteria and treated with 20 units of PlyG for 15 min. RSVF1 was reduced >1.6×10^7-fold, while ATCC 10987 was reduced ~100 fold. Other strains examined were largely resistant, even after a three-hour incubation. PlyG may clearly direct a potent and specific lethal action to the *Bacillus anthracis* cluster, exhibiting a substrate specificity that closely matches the gamma phage host range. Moreover, the capsulated state of several *Bacillus anthracis* strains examined, indicated that capsule does not block access of PlyG to the cell wall.

[168] It was found that PlyG, like most lysins, is a very active enzyme. The addition of 2 units of PlyG to ~1.0×10^4 RSVF1 caused an immediate release of intracellular ATP (measured as light emitted by firefly luciferin/luciferase), which is consistent with a rapid lytic effect. This effect was specific for RSVF1, and was not observed in other isolates tested, therefore suggesting that the ATP release assay is a strong diagnostic tool for g-sensitive bacilli. In a separate kinetic analysis of RSVF1 killing, it found that as little as 2 units of PlyG effected a 1.7×10^4-fold decrease in viability within 20 sec, and sterilization at 2 min. Here the time course of RSVF1 killing in cultures was treated with buffer (r) or 1 unit of PlyG (™). These values are shown as colony forming units per ml of culture determined at each time point. The corresponding OD_{600} determined for the PlyG-treated sample (Î) is shown. The loss of culture optical density lagged behind the

loss in viability, implying that killing by PlyG does not necessarily require extensive cell wall degradation.

[169] *Example 12: Microscopy*

[170] To visually examine the lytic effect phase contrast microscopy of PlyG-treated RSVF1 was used. It was found that the normally filamentous RSVF1 rapidly converts to short rod- and minicell-like forms 30 sec after exposure; nearly complete loss of cytoplasmic material occurs by 15 min, leaving "ghost" cells. Transmission electron microscopy of the rod forms reveals the cytoplasmic membrane bulging from regions of localized cell wall hydrolysis. These structures are usually apparent at polar and septal positions, and rupture to yield a ghost-like form.

[171] *Example 13: In Vivo Lysin Activity*

[172] The lytic effect of PlyG suggested it could be used to kill g-sensitive bacteria in a mouse model of infection. Four- to eight-week old BALB/c female were purchased from Charles River Laboratories and housed at the Laboratory Animal Research Center at The Rockefeller University. Mouse infections were performed as a variation of a previously described procedure. Log phase RSVF1 grown in BHI medium, was pelleted and washed twice in 50 mM $K{\cdot}PO_4$ buffer (pH 7.4). Aliquots of ~1.0×10^6 cells in buffer were injected intraperitoneally (i.p.) into mice in 0.1 ml doses. After 15 min, 0.5 ml of either buffer alone or PlyG in buffer were injected into the peritoneal cavity. Injections of PlyG alone (no bacteria) were also performed to assess toxicity. Mice were monitored for up to 3-4 days, at which time all surviving mice had recovered a normal and unremarkable appearance.

[173] The i.p. injection of some *B. cereus* isolates may induce a rapidly fatal illness similar to experimental anthrax. The injection of ~1.0 x 106 RSVF1 cells into BALB/c mice, killed all subjects in five hours or less. More specifically, mice were injected i.p. with ~1.0×10^6 RSVF1 cfu and treated after 15 min with either phosphate buffer (n=15), 50 U PlyG (n=17), or 150 U PlyG (n=9). As an additional control, mice that were not challenged with bacteria were injected with 50 U PlyG (n=5). The experiment was terminated at 72 hours. Administration of

either 50 U or 150 U to the infected mice was significantly protective compared to the buffer control ($P < 0.0001$). The median survival time for the buffer treated mice was 2 hours. At death, many mice exhibited severe edema at the inoculation site, and hemorrhaging through the eyes and mouth. When PlyG (50 units) was injected i.p. 15 min post-infection, a pronounced therapeutic effect was observed: thirteen of nineteen mice fully recovered, while the remainder survived six to twenty-one hours. When 150 units of PlyG were used, a similar rate of recovery was observed. No toxicity was detected with either the i.p. or i.v. injection of PlyG alone. PlyG does, therefore, rapidly kill g-sensitive bacteria in an infected animal.

[174] The ability of PlyG to degrade germinating spores was examined next. Spores were prepared as described in Mazas, M., Martinez, S., Lopez, M., Alvarez, A. B. & Martin, R. Thermal inactivation of *Bacillus cereus* spores affected by the solutes used to control water activity of the heating medium. Int J Food Microbiol 53, 61-7 (1999). Samples containing 95-99% refractile endospores, as determined by phase contrast microscopy, were stored at 4°C in water. For spore killing experiments, 0.2 ml aliquots of ~2.0×10^8 spores were heat activated at 65°C for 5 min. Samples were pelleted and suspended in 1.0 ml tryptic soy broth (TSB, Difco) containing 100 mM L-alanine (to induce germination) for 5 min at 37°C. The cells were then treated with 1.0 ml of PlyG (10 units) for 5 min at 37°C and plated for enumeration. TSB with L-alanine is a potent inducer of germination for each spore type, converting >99% of each spore type used to heat sensitive forms within 15 min. In the presence of D-alanine, germination frequency was reduced to <10%.

[175] *Example 14: Spore detection*

[176] For spore detection, the spore killing protocol was modified for use with a microluminometer (model 3550i, New Horizons Diagnostics Corp.). Essentially, 0.1 ml of heat-activated spores (65°C, 5 min) were immobilized on a 0.45 mM filter in the 0.4 ml reaction tube. The immobilized spores were washed twice with somatic cell releasing agent and treated with 0.15 ml TSB with 100 mM L-alanine for 5 min at room temperature. Samples were then washed and

treated with 0.15 ml PlyG (2 units) for 5 min at room temperature. 50 ml of a luciferin/luciferase reagent was added for the indicated length of time and a quantitative measure of the resulting light, given as relative light units, was displayed by the luminometer. In the dormant state, the spore's peptidoglycan, or cortex, is protected from lysozymes and amidases by a proteinaceous coat. However, within 10 min of inducing germination, coat porosity increases as it begins to degrade; it was reasoned that subjacent peptidoglycan may be rendered susceptible to PlyG.

[177] To evaluate this, spores were prepared from RSVF1, closely related *B. cereus* (ATCC 14579) and *B. thuringiensis* (ATCC 33679) strains, and *B. subtilis*. Aliquots of ~108 heat activated spores were induced to germinate for 5 min and then treated with PlyG (10 units) for 5 minutes. Resulting spore viability was compared to that of spores treated with D-alanine, a germination inhibitor. While all D-alanine-treated spore samples were PlyG-resistant, only RSVF1 was sensitive after germination in the presence of L-alanine., showing a dramatic decrease in viability of log10 3.9. A sporocidal action, therefore, occurs rapidly after the induction of germination, when PlyG may likely access the cortex. In light of the thickness of the cortex, the rapid PlyG effect suggests a subtle alteration impairing spore outgrowth, rather than a massive degradative action.

[178] The ability of PlyG to kill germinating spores was exploited to develop a rapid and specific system for detecting g-sensitive spores using a hand-held luminometer. Spores were immobilized or placed on filters or in cuvettes (in a solution) and incubated in at least one 5 min round with at least one germinant and PlyG (2 units). The temperature at which the incubation took place was from room temperature to 60 degrees Centigrade. The spores could be incubated first in germinant and then in PlyG or with the germinant and PlyG together. The phage associated lytic enzyme does not have to be PlyG but are preferably specific for the spore being tested. The release of ATP from degrading spores was then measured as a light "flash" emitted in the presence of a luciferin/luciferase reagent. ATP released from PlyG-treated germinating spores was assessed in the presence of luciferin/luciferase. 2.5 x 103 RSVF1 spores were induced to

germinate with L-alanine and treated with 2 units of PlyG. The PlyG-mediated flash was measured. Germinating spores of Bc 14579, Bt 33679, and *B. subtilis* showed no activity, demonstrating the expected recognition specificity of PlyG. Not surprisingly, when spore preparations were mixed, only the combination containing RSVF1 yielded a light signal Samples containing 2.5 x 10^3 spores of Bc 14579, Bt 33679, and *B. subtilis* with (RSVF+ mix) or without (RSVF1- mix) RSVF1 were induced to germinate in L-alanine. The intensity of luminescence after PlyG treatment (2 units) was measured. The sensitivity of our system was examined using samples containing as few as ~100 spores. Rather than an immediate light flash, an RSVF1 signal was observed after 60 min incubation in the presence of PlyG and the luciferin/luciferase reagent. This signal is consistent with a low-level "glow," and is consistent with the low levels of ATP likely being released. No glow was detected in the presence of other germinating spore types, and is, therefore, specific to the g-sensitive spores. This sensitivity, combined with the specificity, rapidness, and highly portable nature of our detection method, suggests applications in monitoring both domestic and battlefield use of *Bacillus anthracis* as a biological weapon. This technique may be used to identify the presence of spores from other bacterial species using bacteriophage lysins specific for those species.

[179] The phage associated enzyme used to lyse the *Bacillus anthracis* spores may be a lytic enzyme, chimeric lytic enzymes, shuffled lytic enzymes, and combinations thereof. The phage associated lytic enzyme, and its altered forms, may be the PLY gamma enzyme, or another phage associated lytic enzyme specific for *Bacillus anthracis*.

[180] A holin protein may also be used to assist in the lysing of the germinating spores. The holin protein may be unaltered, chimeric, shuffled, or may be combinations, thereof.

[181] The nature of the luminometer that may be used for the detection of ATP, and its method of use is found and described in U.S. Pat. 6,395,504 (herein incorporated by reference).

[182] *Example 15: Mutagenesis and screening for resistance*

[183] Spontaneous lysin resistance was initially examined as described (Loeffler, et al.) by repeated exposure to PlyG at low concentrations on agar plates or to increasing concentrations in liquid assays. No resistance was detected.

[184] To determine if spontaneous resistance was at all possible, chemically mutagenized cells were examined. Log phase RSVF1 was treated for 4 hours with methanesulfonic acid ethyl ester (EMS) at a concentration of 150 mM, resulting in 90% killing. The cells were then washed with BHI and grown 3 h (three cell doublings) prior to freezing at -70°C. The efficiency of mutagenesis was estimated by the frequency of mutations giving resistance to 150 mg ml^{-1} streptomycin ($strep^R$) or to 3.5 mg ml^{-1} novobiocin (nov^R). The spontaneous frequencies in non-mutagenized cultures were 2.4 $x10^{-9}$ for $strep^R$ and 4.0 x 10^{-10} for $strep^R$; for EMS treated RSVF1 the frequencies were 2.1 $x10^{-6}$ for $strep^R$ and 4.3 $x10^{-6}$ for $strep^R$. For screening, frozen mutagenized cells were then thawed, washed in BHI, and grown for 30 min at 30°C. One milliliter aliquots (~1.0 $x10^8$ cells) were incubated with PlyG for 30 min at 37°C, washed, and either plated or incubated overnight in BHI. Colonies arising from the plated cells were picked and evaluated for resistance to 20 units of PlyG in the spectrophotometric lysin assay. For the overnight BHI cultures, log phase cells were established and ultimately treated again with PlyG as before; this was repeated for 4 successive days. In one set of experiments, 20 units of PlyG was used for each treatment, while in another 0.05 units was used and followed by serial 10-fold higher doses on following days. Bacteria were plated after each treatment, and later examined for resistance to 20 units of PlyG in the spectrophotometric lysin assay. No resistance was detected.